nature

www.nature.com/nature

Vol 446 | Issue no. 7137 | 12 April 2007

Climate change, here and now

Poor nations need the data that show what is already happening to their climate, as well as the resources with which to adapt to change.

Critics are always ready to accuse the Intergovernmental Panel on Climate Change (IPCC) of exaggeration. The content of its sobering 6 April report on climate-change impacts, adaptation and vulnerability (see page 706) offers little support to such criticism. But the panel is, from time to time, guilty of almost absurd understatement.

While discussing the encouraging growth in real data applicable to the study of climate change, Working Group II of the IPCC points with regret to a "notable lack of geographic balance in data and literature on observed [climate] changes, with marked scarcity in developing countries". A few pages later, a figure in the report reveals that of 28,671 "significant observed changes in biological systems" from around the globe of which the report made use, 28,115 originated from Europe. Just two were from Africa.

This disparity is all the more alarming because, as the report makes clear, it is in Africa and other parts of the developing world that such data are most sorely needed. They are required not as proof of the global reality of change — that debate is over — but as guidance to policies and interventions that are needed, not in a far-off world of melted ice caps, but right now. The data are needed so that policymakers can know what is happening to crops, to river flow, to soil moisture, and make appropriate use of that information.

In terms of its scope, anthropogenic climate change is unlike any problem previously faced by humanity. Its effects are already felt worldwide and will last for centuries. Solutions to its primary cause — the release of carbon dioxide from the burning of fossil fuels — are measured in terms of trillions of dollars of investment taking place over decades. Yet it also has more immediate, short-term implications. Working Group II predicts with high confidence that, by 2020, between 75 million and 220 million Africans will suffer from increased water stress due to climate change. In the same period, in some African countries, yields from rain-fed agriculture could be halved.

The qualitative sense in which 2020 can be clearly set down as the 'short term' is that predictions made in the report for that date are not conditional. They do not depend on action or inaction on greenhouse-gas emissions between now and then. Rather, they are immune to anything being done about emissions at any point. If the industrial nations were to start slashing their emissions on 1 May of this year, the prognosis would be the same, because the degree of change to be expected through to 2020 is already pretty well fixed by the current state of the climate system.

This is not for a moment to say that there is no point in reducing emissions. That must remain a central objective in attempting to get the planet's climate under control. But Working Group II's report makes clear that this is not enough.

"Developing countries need assistance to help them adapt to effects of climate change that are already on the way."

Developing countries also need assistance that will help them adapt to effects of climate change that are already on the way. Although some of this may be focused on climate-specific approaches (see page 716), most of it can best be achieved through conventional aid programmes aimed at economic development, agricultural robustness and primary health care. Better off, healthier people are people more likely to have the resources needed to adapt.

Both developed and developing countries also need far better data about what is happening in their territories. Entire areas of study, such as the retreat of permafrost (see page 718), remain bereft of appropriate data sets.

And energy sources and patterns of use must shift to lessen the ultimate extent of climate change. But the world also needs to take action right now against the harm that is already being done. The IPCC's latest report provides yet another reason to seek development pathways out of the pitiless poverty in which far too much of humanity is trapped. ■

Science without borders

Researchers should push for rule changes to make Europe work as one.

If European science is to prosper, the barriers that prevent seamless interactions between scientists in different nations need to come down. Last week, the European Commission published a 'green paper' on the future of the European Research Area, the entity it created to improve such interactions. Scientists must now engage in the consultation process that will follow from this document and so help resolve the problems that currently constrain 'cross-border' science.

The European Research Area is a somewhat nebulous concept, most readily described as the highly fragmented arena within which European Union scientists work, in both public and private sectors. But it is a concept that matters, in determining how easily a European researcher can operate across national borders.

Current deficiencies in the way the area works are most apparent when they are personal. One German biology professor, for example, was recently courted by a university in the Netherlands. Aware of the advantages the prestigious post held for his research ambitions, the professor was sufficiently enthusiastic to accept a small salary cut. But negotiations collapsed because Dutch regulations made it impossible for him to bring his German pension to the new position.

On a broader level, those seeking to create expensive items of

scientific infrastructure to serve the whole continent, such as a biobank of gene or tissue samples, are often dismayed to find that there is no appropriate source of funding — no matter how useful the project. Scientists in Europe also find that the management of intellectual-property rights varies between the member states. Scores of issues such as these make it hard for researchers to properly exploit the scale of the European Research Area.

Overcoming such obstacles is rightly seen as important in the European Union's push to become more scientifically competitive with the United States, where scientists already enjoy the advantages of ready interaction with a vast array of colleagues in a nation of 300 million people.

The green paper released by the European Commission on 4 April outlines the existing problems and asks for ideas from interested parties, including scientists, on what should be done to fix them. The consultation process will include a questionnaire that will appear on the commission's website (http://ec.europa.eu/research/era) from 1 May until August, and a conference in Portugal this autumn. Early next year, the commission will use this feedback to help it draw up decrees or legislation that it thinks will help strengthen the European Research Area.

But the commission does not have the clout to implement such reforms on its own. Its main political master, the Council of Ministers, committed itself in 2000 to improving competitivity in research and innovation by 2010 — by facilitating the mobility of researchers, for example. But the member states whose leaders make up the council have not yet implemented the changes in their home countries that are needed for the European Research Area to function effectively.

The European Commission, whose Framework research programmes still account for only about one-twentieth of the member states' total spending on research, can do little more than encourage national governments to realize that far more cooperation at the European level will benefit them all. Perhaps the most sensitive issue in this regard is to get more national tax revenue to be pooled for genuinely European projects.

"It is important that the European Commission gains the explicit support of both industrial and academic scientists in its long march towards European research unity."

None of these problems will be solved overnight. But it is important that the commission gains the explicit support of both industrial and academic scientists in its long march towards European research unity. The consultation will allow individual researchers to put on the record the cross-border issues that confront them in their working lives. As many as possible should fill in the questionnaire and make their voices heard. ■

When employees attack

Government scientists should be able to comment publicly — within reason.

Badmouthing one's government is a fashionable pastime in some parts of the world. Many US climatologists, even those who receive federal funding, have grave reservations about the White House's continued neglect of international climate agreements, and they aren't shy about saying so. In Britain, meanwhile, scientists as well as political analysts have been quick to criticize the government's plan to spend billions on renewing the national fleet of nuclear-weapons submarines.

Roll those two examples together, and transplant them into a society where freedom of speech is often seen as being under pressure from several directions, and you get the case of Claudio Mendoza. Until recently the head of a government physics laboratory in Venezuela, Mendoza has been demoted after making sarcastic comments about the government over what he regards as its tendency to ignore scientists and their advice (see page 711).

What infuriated Mendoza's paymasters most was probably his suggestion — made in a newspaper article promoting a play about nuclear weapons — that president Hugo Chávez might want to pursue a nuclear-weapons programme and that, if he did so, he was liable to fail because of this alleged disdain for expert advice.

Mendoza's comments were not made in any official capacity (his article was signed, with no affiliation given), raising the fraught question of whether senior government scientists should be free to make disparaging public comments about the state institutions that they serve, when they are away from work.

On a facile level, this is a disagreement about whether it is acceptable for someone to be fired because their bosses can't take a joke. In many countries, acerbic comments about the machinations of politics are a valid and effective mode of public discourse.

But, of course, a line has to be drawn somewhere. It is hard to escape the feeling that, in this case, it has been drawn in the wrong place. Many civil servants in other countries might expect a dressing-down if they behaved in this way, but might justifiably argue that they have a right to express a grievance. The message coming from Mendoza's bosses within the Venezuelan national research institute is an unsavoury one. His removal from a management position implies that someone who voices contrary opinions is not fit to be a lab head. What's more, Mendoza has been warned that he had better clam up if he doesn't want to lose his job altogether.

The play that Mendoza was writing about was Michael Frayn's *Copenhagen*, the international hit that deals with a crucial 1941 meeting between Niels Bohr and Werner Heisenberg, and their struggle to comprehend the feasibility and consequences of developing nuclear weapons during the Second World War (see *Nature* **394,** 735; 1998).

One of the reasons for the play's success was general interest in what physicists of Bohr's generation thought about the issues surrounding nuclear weapons. Of course, these thoughts only became public some time after the United States had built and used the bomb. But times have moved on, and people in Caracas, as elsewhere, would benefit if their scientists were be able to participate openly in public debate on nuclear policy. ■

RESEARCH HIGHLIGHTS

Arid America

Science doi:10.1126/science.1139601 (2007)
The great American drought that spawned the Dust Bowl of the 1930s may become the new climatic norm for much of southwestern North America.

Researchers led by Richard Seager at Columbia University in New York have developed a model projecting the impact of anthropogenic warming on the climate and hydrology of the already arid regions of the United States and northern Mexico. Using 19 climate models, the team projects an increase in aridity over the next several decades comparable with that of the drought that marked America's Dirty Thirties (pictured). The findings are also applicable to other subtropical areas.

BETTMANN/CORBIS

GENETICS

The long and short of it

Genome Res. doi:10.1101/gr.6036807 (2007)
Do long RNA molecules that do not encode proteins have any function? A team at the University of Oxford, UK, presents evidence that suggests they do.

Gerton Lunter and his colleagues compared the genetic sequences of more than 3,000 long noncoding RNAs from the mouse with the equivalent genomic locations in humans and rats. They identified stretches of DNA related to the RNAs (RNA libraries have not been compiled for humans or rats yet) that were conserved across the species. These included the promoter regions that prime transcription of DNA into RNA. This implies that the long RNAs have some biological function that is worth preserving — although what that function is remains a mystery.

MATERIALS CHEMISTRY

Emerging patterns

Macromolecules **40,** 1594–1597 (2007)
Chemists have chanced upon a way to deposit metal atoms at the interfaces between different regions of a 'block co-polymer' — molecules that contain distinct segments of different types of polymer. The resulting patterns of atoms could be exploited in electronic devices.

A team led by Mitchell Winnik of the University of Toronto, Canada, and Ian Manners of the University of Bristol, UK, noticed that ruthenium tetroxide, a stain used for electron microscopy, concentrates at the interfaces in structures self-assembled from a certain diblock co-polymer. This unusual phenomenon probably results from the stain dissolving in one polymer type, then diffusing towards the other. It reacts with iron-containing groups in the second polymer and deposits ruthenium atoms.

ACOUSTICS

Unsound judgement

J. Acoust. Soc. Am. **121,** 2384–2395 (2007)
Violin-makers don't pick the materials shown by scientific measurements to be acoustically best, according to new research.

Christoph Buksnowitz of the University of Natural Resources and Applied Life Sciences in Vienna, Austria, and his co-workers asked 14 leading luthiers to grade 84 pieces of Norway spruce — a favourite material for violin soundboards — according to quality.

The results showed little correlation with laboratory tests of the materials' acoustic properties, but seemed instead to be guided by the wood's appearance. That's not totally damning, the researchers say, because a violin's performance may depend as much on what the luthier does with the wood as on its intrinsic properties. But it does suggest that violin-makers would benefit from a scientific definition of material quality.

A. WINKLER/REUTERS/CORBIS

CLIMATE SCIENCE

Mushrooming harvest

Science **316,** 71 (2007)
Climate change could make the autumnal fungus foray a year-round event, say Alan Gange of Royal Holloway, University of London, and his colleagues.

Analysing changes in fruiting patterns of more than 300 species of British fungus between 1950 and 2005, the team found that the autumn mushroom season has grown from about 33 days to nearly 75 days. And some species, such as the sulphur tuft (*Hypholoma fasciculare*) are fruiting in both spring and autumn.

The extended autumn season is due to warmer summers, wetter autumns and fewer frosts. The additional spring season is probably triggered by warmer weather in late winter and early spring.

IMMUNOLOGY

RAGE management

Nature Immunol. doi:10.1038/ni1457 (2007)
Researchers have identified one way in which the immune system turns against the body's own proteins to trigger inflammation and cause the common autoimmune disease systemic lupus erythematosus.

Anthony Coyle of MedImmune in Gaithersburg, Maryland, and his colleagues studied HMGB1, a DNA-packaging protein that is released from dying cells. They

show that it forms part of a complex that is recognized by the cell-surface receptor RAGE and taken up by immune cells. Once inside the cells, the complex stimulates Toll-like receptor 9, which normally causes inflammation by recognizing DNA from invading bacteria and viruses.

Antibodies that block human HMGB1 or RAGE cut the production of inflammatory signals from cells *in vitro*, suggesting a means of therapy.

NEUROSCIENCE

Seizure stopper

J. Neurosci. **27,** 3618–3625 (2007)

The low-carbohydrate 'ketogenic' diet used by some people to control their epilepsy may work by boosting the activity of ATP-sensitive potassium (K_{ATP}) channels, scientists propose. This would inhibit excessive electrical activity in the brain.

The ketogenic diet causes tissues to use ketone bodies (the by-products of fat breakdown) instead of sugar as fuel. Gary Yellen and his colleagues at Harvard Medical School in Boston, Massachusetts, report that ketone bodies reduce electrical activity in rodent brain slices, particularly in cells with high firing rates. The effect is lost when K_{ATP} channels are pharmacologically blocked, and does not occur in the brains of transgenic mice lacking K_{ATP} channels.

METROLOGY

Current affairs

Nature Phys. doi:10.1038/nphys582 (2007)

A new way to define the ampere, the unit of electrical current, is in prospect thanks to a microelectronic device built by Mark Blumenthal of the University of Cambridge and the National Physical Laboratory in Teddington, UK, and his co-workers.

At present, the ampere is defined as the current required to create a particular force between two parallel wires a metre apart. Physicists would rather define it as the passage of a particular number of electrons per second past a fixed point, but they first need a quantum device that can reliably create a current of known electron flow to serve as a standard.

Blumenthal and his colleagues have built a 'pump' that operates at gigahertz frequencies — transmitting around 3.4 billion electrons per second. This equates to about half a nanoamp of current, close to that required for a reliable standard.

TAXONOMY

Death of an order

Biol. Lett. doi:10.1098/rsbl.2007.0102 (2007)

Termites should be classed with the cockroaches in the order Blattodea, and not granted their own order Isoptera, say researchers who have carried out the most comprehensive molecular study of these creatures' relationships so far.

It was known that the two groups were closely related. In fact, the termites' family tree falls entirely within that of the cockroaches, say Paul Eggleton of the Natural History Museum in London, UK, and his colleagues.

D. HOSKING/FLPA

The researchers sequenced five genes from 107 species in the group Dictyoptera, which consists of termites, cockroaches and mantids. The termites' closest roach relative is the woodroach *Cryptocercus*, suggesting that wood eating was a key innovation on the termites' road to sociality. The picture (above) shows a termite mound.

CHEMISTRY

Running rings

J. Am. Chem. Soc. **129,** 4148–4149 (2007)

Two fiendishly complex molecules isolated from the clubmoss *Lycopodium complanatum* have succumbed to synthesis in the lab, six years after the first of the pair was described.

The challenge in synthesizing the compounds, known as (+)-Lyconadin A and (–)-Lyconadin B, lay in constructing the five interlinked rings that form the molecules' core. Four of these rings are loops of six atoms, and one is a five-membered ring that contains four chiral centres, meaning the groups around them must have a particular symmetry.

The strategy that Douglas Beshore and Amos Smith of the University of Pennsylvania in Philadelphia devised to build this structure includes one step that creates a three-ring system from a molecule that begins with only one ring.

JOURNAL CLUB

Kornelia Polyak
Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts

A geneticist rebuts criticism of cancer genome projects.

What do you learn if you sequence 13,000 genes in 11 breast and 11 colorectal cancer samples? The question taps into an intense debate about how best to identify genes relevant to human cancer.

Last year, researchers reported the results of a survey such as the one described (T. Sjöblom *et al. Science* **314,** 268–274; 2006). They found that each tumour contains, on average, 90 mutant genes — an unexpectedly high number. They also defined mutation spectra that were specific to colon and breast tumours, including the intriguing observation that the DNA letter sequence CG was swapped for GC at high frequency in breast tumours. This could be due to an uncharacterized DNA repair defect or differential carcinogen exposure.

I consider this report a step towards answering key questions in cancer biology, such as how many genes are mutated in cancer, how many mutations are required for cancer, and whether accumulation of genetic alterations in cancer cells drives tumour progression.

But others disagree. Many labs see large-scale sequencing of cancer genomes as unfocused and expensive fishing experiments. I have been doing genomics experiments since the dawn of this era, and have often faced this criticism.

But just this one study has identified more genes mutated in human cancer than thousands of investigators have found over past decades. And another recent, large-scale sequencing project pinpointed close to 120 mutant kinase enzymes that may have a role in human cancers (C. Greenman *et al. Nature* **446,** 153–158; 2007).

Both cases show that the outcome of unbiased, genome-wide studies may not be what we expect, which is exactly why they're worth doing.

SPECIAL REPORT

Climate takes aim

Attention is now turning to the developing world, where those least equipped to handle it will bear the brunt of global warming. **Michael Hopkin** reports.

BRUSSELS

The Intergovernmental Panel on Climate Change (IPCC) is not known as a bearer of good news. In February, it reported that human activities are almost certainly causing the planet to warm (see *Nature* **445**, 578–579; 2007). On 6 April in Brussels, Belgium, it delivered an even more sobering message: that billions of the world's poorest citizens are at risk of hardship and disease as a result of climate change.

Attention is now shifting from arguments over whether the world is warming to what should be done about it. And all six-billion-plus on the planet should be concerned, the IPCC's report implies. The people most vulnerable will be those who live at or near sea level, often crowded into cities along the coast. But drought, disease and extreme weather events will also become more frequent around the world, threatening the lives and livelihoods of countless more.

"No one will escape the impacts of a warming planet," says Patricia Romero Lankao, a sociologist at the National Center for Atmospheric Research in Boulder, Colorado, and an author of the report.

What's more, many of the effects of climate change are already evident in physical and biological systems, the report says (see map). Regional climate changes are affecting natural systems on every continent, with the Arctic, sub-Saharan Africa, and Asian mega-deltas among the worst affected.

The *Climate Change Impacts, Adaptation and Vulnerability* report is the second instalment of the IPCC's Fourth Assessment, a summary of the current state of knowledge about climate change. The third assessment was published in 2001. Crucially, this report is the first to link actual data on how natural systems are responding to the amount of warming they have experienced. "For the first time we are no longer arm-waving with models," says Martin Parry, co-chair of the IPCC's Working Group II.

Authors compiled more than 29,000 data sets, on everything from glaciers to the timing of spring foliage, and compared the trends with the amount of regional warming observed in each area since 1970. In more than 90% of cases, the changes in natural systems were consistent with predictions of how they would behave in a warming world. "Regional changes are happening on a global scale," says Parry, although more data are needed from the developing world.

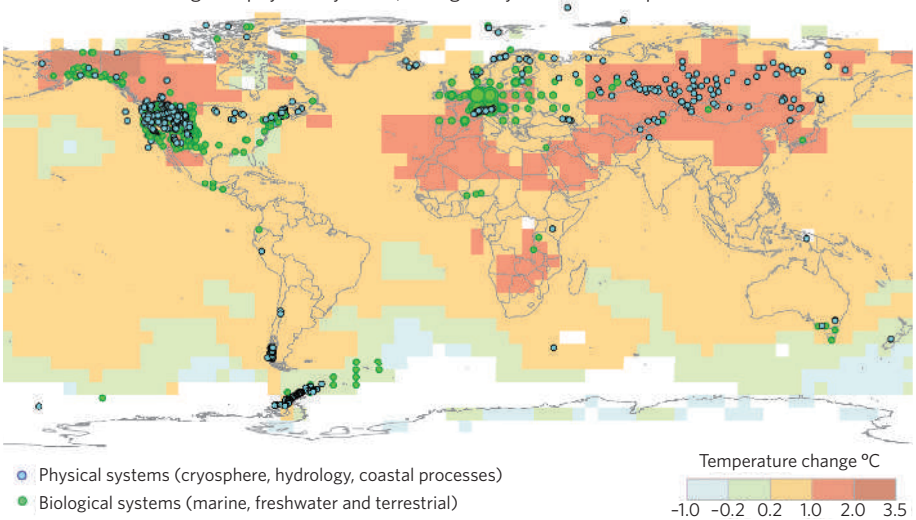
The report was also the first major attempt to predict what physical changes might take place under a range of future warming scenarios. At high latitudes such as in northern Europe, crop yields are expected to flourish under modest climate warming. But rises of greater than about 2 °C look set to diminish agricultural productivity the world over (see 'What's new').

By evaluating which systems are the most likely to be altered by future climate change, the report's authors have tried to predict the knock-on effects for humans. Most at risk are the estimated 100 million people who live within a metre of sea level. Meanwhile, as many as 250 million Africans are predicted to be at risk of water shortages, and drought will also damage agriculture at low latitudes. "We have 500 million people hungry today," says Parry. "We're gonna get a lot more."

In terms of health, vector-borne diseases such as tick-borne encephalitis and dengue fever are expected to increase as insects expand their range. So too will extreme weather events such as the 2003 heatwave that killed thousands of people in Western Europe.

The report builds on that of the IPCC's first working group, which announced in February that it had 'very high confidence' that global warming can be attributed to human activities. But whereas the report from the first working group was carefully couched in language of confidence and likelihood, many such phrases are absent from the policy-makers' summary of the second working group's report. Before the launch in Brussels, the report's lead authors stayed up all night deliberating the finer points of the policy-makers' summary, which boils more than 1,500 pages down to a mere 23. Several delegates from the research community

Observed changes in physical systems, biological systems and temperature for 1970–2004



SOURCE: IPCC

NASA

CLIMATE CHANGE
Find all our coverage of the changing climate online.
www.nature.com/news/infocus/climatechange.html

FLPA

At high latitudes, crop yields may improve with modest climate warming.

What's new?

Compared with the last report of the Intergovernmental Panel on Climate Change published in 2001, the latest version adds some details about the effects that can be expected from climate change. These include:

Some 20–30% of animal species — at least those studied so far — could be at risk of extinction if global temperatures rise more than 1.5–2.5 °C.

Fisheries in large lakes in Africa could be adversely affected not only by overfishing but also by rising water temperatures.

Glaciers melting in the Himalayas could reduce the amount of water available for people to drink, as well as destabilizing slopes and lessening river runoff.

More drought and fire could place croplands and forests in southern and eastern Australia at risk.

Agricultural land in Latin America could turn to desert and grow saltier.

On small islands, invasive species could arrive sooner as temperatures rise.

stormed out during the negotiations, albeit to return later. China and Saudi Arabia, in particular, were said to have put pressure on to soften the language.

Cynthia Rosenzweig of NASA, a lead author of the report and one of those who left the negotiating session, says that the differences over the confidence statements were because many of the government delegates wanted the statements to be ditched. "But we wanted governments to know that we felt strongly that the 'very high confidence' statement was justified," she says. Statements of confidence remain in many phrases of the summary, and the altered phrases now reflect the evidence used to support the findings. One passage reads, for instance, that "Observational evidence from all continents and most oceans shows that many natural systems are being affected by regional climate changes, particularly temperature increases." Rosenzweig says: "We have a direct statement now, which in some ways implies a confidence level." Removing qualifications, she adds, "has nothing to do with climate evangelism".

Although the report's predictions are certainly serious, Parry denies that they are dire. "It is a good, tight process, and the real key is that governments have to buy into it — it is an intergovernmental panel," he says.

That government buy-in should take the form of both adaptation and mitigation — the two buzzwords of how to deal with climate change. Although the second working group's brief was to examine only one of these approaches — adaptational measures, such as improved healthcare or flood defences — the report also included a chapter on combined approaches that involve both adaptation and mitigation, such as carbon-reducing policies. The third working group of the IPCC is due to release its report, specifically on mitigation, in Thailand on 4 May.

"No one will escape the impacts of a warming planet."

Adaptation is the key to the next few decades, says Saleemul Huq of the International Institute for Environment and Development in London, one of the report's authors. The report says that even if no more carbon is put into the atmosphere, average warming of 0.6 °C can still be expected over the rest of the century. "Adaptation is the only option in the short term," Huq says. "However, as we extend to the longer term — the next ten or fifteen decades — the only solution for that is to do mitigation now. If we fail to do either of them now we will suffer."

Ivo de Boer, executive secretary of the United Nations Framework Convention on Climate Change, which governs the Kyoto Protocol regulating greenhouse-gas emissions, hopes that the report will provide impetus for negotiations for a climate treaty to replace Kyoto when it runs out in 2012. "I hope it will make it clear to politicians that time is running out," he says.

One of the cruel ironies is that among the few set to gain, at least in the short term, from the agricultural benefits conferred by climate warming are those with the highest greenhouse-gas emissions. And yet the central message of the report is that climate change is likely to hit hardest those who can do least to defend themselves.

Nevertheless, the risks are not confined to poor countries, says IPCC chair Rajendra Pachauri. "It is the poorest of the poor, even in richer societies, who will be affected the most significantly," he says. "People who are poor are least equipped to be able to adapt to the consequences of climate change." ■

Michael Hopkin

Dinosaur prints lead to crediting row

An article on Croatian dinosaur tracks that help to reveal the tectonic movements responsible for the creation of the Balkans has set off a modern-day continental clash. The situation raises questions about how far palaeontologists should have the right to prevent others from publishing on sites of interest that they discover.

The huge sauropod footprints, dating from about 95 million years ago, were discovered in 2004 on the island of Hvar by a team from the Croatian Natural History Museum in Zagreb, led by Michael Caldwell of the University of Alberta in Edmonton, Canada. The museum announced the discovery of the site to the press. But before the researchers could publish their study in the scientific literature, a competing group from the University of Zagreb went to the site and then published an article reporting the tracks. The article did not credit the discovery team, and an angry Caldwell has demanded its retraction.

The disputed article was published last December in the journal *Cretaceous Research* (A. Mezga *et al. Cretac. Res.* **27,** 735–742; 2006). The lead author was palaeontologist Aleksandar Mezga, working with three Croatian colleagues and a Swiss researcher.

A. MEZGA

M. CALDWELL

***Faux pas*? Dinosaur footprints discovered by Michael Caldwell (inset) and his team have resulted in discord.**

Caldwell says he believes that the publication amounts to "intellectual theft", because it leaves his team unable to publish the discovery as its own. Caldwell's colleagues in Croatia declined to speak to *Nature*, but e-mails from them provided by Caldwell suggest they support his interpretation of events. For example, when he learned of the article in February, Jakov Radovcic, a Croatian sedimentologist working with Caldwell, wrote: "I was really shocked. I was not aware of the... behaviour of my colleagues from Zagreb."

Mezga, however, insists that he and his team have done nothing wrong. "Every single word... is our intellectual property," he told *Nature*. Because nothing had been published on the study site, Mezga says, he considered it open for reporting, although he acknowledges that he never contacted the discovery team to discuss the issue.

David Batten, a researcher at the University of Manchester, UK, and editor-in-chief of *Cretaceous Research* at the time the article was published, told *Nature* that the journal is considering Caldwell's request for a retraction. Such a retraction would be "very unusual", says Catherine Badgely of the University of Michigan in Ann Arbor and president of the US-based Society of Vertebrate Paleontology. Documented instances of misappropriation of palaeontological sites are rare, surfacing about every five years, Badgely notes. But she welcomes discussion of the issue: "It is good this is being brought out."

In this case, the tracks of a long-necked, plant-eating sauropod known as a titanosaur provide insight into massive tectonic shifts that occurred around 100 million years ago. The land mass that is now Africa was then shifting north against the Eurasian continent, raising and lowering geological structures across what have become the Balkan nations.

The dinosaur tracks are the oldest to have been found in that region. But they also represent the most recent signs of life on a geological plain called the Adriatic–Dinaric carbonate platform, before it sank beneath the sea 94 million years ago. And they provide an important clue about what other land the platform was connected to at the time.

Caldwell and his colleagues were exploring Hvar's rocky shoreline when they happened upon the 10 pad marks — each about 30 centimetres long — in a 65-square-metre area of ground.

It is common for museums to announce such discoveries, and the Hvar site drew wide news coverage. The second Zagreb team started working on the site within a couple of months. They brought with them Christian Meyer, an authority on dinosaur tracks and director of the Natural History Museum in Basel, Switzerland. When contacted by *Nature*, Meyer declined to comment on the team's rights to the site, referring inquiries to his Croatian colleagues. But in an e-mail sent to Caldwell in February, he said his colleagues had told him no one else was studying the tracks.

Actually, Caldwell's team was analysing tiny plankton fossils to pinpoint the date of the tracks. Those studies were delayed when samples being shipped from Croatia to Edmonton were misplaced for six months by an airline.

Italian palaeontologist Fabio Dalla Vecchia was one of the reviewers of the *Cretaceous Research* article. He says he pointed out to the journal and the authors the need to fully credit the tracks' discoverers. But this did not happen (the journal admits receiving the request, but says that it was misunderstood). "I had nasty experiences in Croatia," adds Dalla Vecchia. "The groups fight each other there."

Caldwell says he is sad that things didn't work out differently. "This could have been such a good thing for Croatian research." ■

Rex Dalton

DINOSAURS IN FOCUS
All our coverage of fossil finds is gathered online.
www.nature.com/news/infocus/dinosaurs.html

Physicists question model of the Universe

Cosmologists gathered in London last month to voice their concerns over the current 'standard' model of the Universe. "There is a sense of desperation," says participant Douglas Scott, from the University of British Columbia in Vancouver. "The standard model is horribly ugly, but the data support it."

For most in the field, that desperation stems from two unexplained ingredients in the standard model: dark matter, thought to help ordinary matter clump together to form galaxies; and dark energy, invoked to explain the observation that the expansion of the universe is accelerating.

But physicists are increasingly questioning the model itself. Richard Lieu, from the University of Alabama in Huntsville, is one of those with doubts. He says he organized the meeting, at Imperial College London, "for scientists to come forward and present any misgivings they may have about the standard model", although he invited speakers from all sides of the debate.

A broad consensus over the standard model has emerged over the past decade. It is based on diverse lines of evidence, such as measurements of the radiation left over from the early Universe (known as the cosmic microwave background), the distribution of galaxies and the brightness of distant supernovae.

"The standard model is horribly ugly."

Scientists at the meeting were asked to re-examine assumptions made when these observations were analysed. For example, Subir Sarkar, from the University of Oxford, UK, argued that data from the Wilkinson Microwave Anistropy Probe (WMAP) — the NASA satellite that measured the cosmic microwave background with unprecedented precision — make sense without dark energy if density perturbations in the early Universe had a different pattern to that usually assumed, and if our patch of the Universe differs from other parts. This explanation might seem more complicated than the standard model, says Sarkar, but he counters that invoking dark energy is "a profound problem from the viewpoint of fundamental physics".

Attendees were also asked to question the actual data. Speakers pointed out oddities in WMAP's maps of the microwave sky, such as the alignment of patterns on different scales along what has been cheekily dubbed an 'axis of evil', and the presence of more variation in the southern hemisphere than in the northern. Such unlikely effects could just be statistical quirks, but they seed uncertainty about how the measurements are being interpreted.

For the time being, proponents of the standard model say that they have yet to see a compelling theoretical alternative, but encourage those people who pursue them. "I doubt that anyone left that meeting convinced about some new idea," says Scott. "But having one's mind opened a little bit is healthy." ■

Jenny Hogan

Venezuelan free-speech row goes nuclear

Freedom-of-speech groups have expressed concern at the treatment of a prominent Venezuelan physicist who has been fired as head of a government research lab after poking fun at the government over nuclear policy issues.

Claudio Mendoza was stripped of his position as head of a computational-physics lab in the Venezuelan Institute of Scientific Research (IVIC) in Caracas because of comments he made in an article written to promote a science-related play. He sarcastically suggested that Venezuelans should not worry about their country's growing alliance with 'rogue' nuclear states such as Iran, because Venezuelan officials do not listen to experts and so would not be able to develop nuclear technology anyway.

J. CASTILLO

"I am just basically trying to survive as a researcher."
— Claudio Mendoza

Although Mendoza is still a researcher in the lab, his dismissal as head after 10 years raises fears that his right to free speech has been infringed, says Juan Carlos Gallardo, chair of the American Physical Society's Committee on International Freedom of Scientists. The committee has written to Venezuelan officials to request details of the case. Although no other scientists there have reported similar harassment, the government has been accused of waging a campaign against freedom of speech in the media, and the fear is that similar repression is now extending to the research community. Gallardo has pledged to monitor the situation and take further action if Mendoza is sacked outright.

Mendoza says he has been accused of treason, even though his comments were meant to be witty and he was not writing in an official capacity. His remarks were published on 13 September 2006 in an article to publicize a production of *Copenhagen* by British playwright Michael Frayn. The play dramatizes a discussion between physicists Neils Bohr and Werner Heisenberg about the feasibility of developing nuclear weapons. Addressing fears that Venezuela might seek to join the nuclear club, Mendoza wrote: "Here bridges are built without engineers, diagnoses are made without doctors, oil is refined without petroleum experts, one can teach without being a teacher, you can govern without being a statesman. We will therefore explode nuclear energy while ignoring the physicists."

But it seems that nuclear policy is no joking matter. Although Venezuela has no nuclear programme of its own, it has significant reserves of uranium ore, and in 2005 Venezuela announced that it would join forces with Iran to develop domestic nuclear power. Venezuela is also thought to have endorsed Iran's controversial uranium-enrichment programme, although without a seat on the UN Security Council, it was unable to influence the council's unanimous vote in December 2006 to ban the project.

Four days after the article was published, IVIC's board of directors removed Mendoza as lab head, and gave him 30 days to provide evidence of his apparent insinuation that Venezuela might be planning to enrich uranium. Mendoza submitted a dossier of newspaper articles but this was rejected as sufficient proof. When asked to retract his article, he refused.

The article was "the last drop" in a series of altercations in which Mendoza has criticized his paymasters, says IVIC director Máximo García Sucre. In 2003, for example, Mendoza complained that the government was not giving enough financial support to IVIC — a claim denied by IVIC directors (see *Nature* **422,** 257; 2003).

"He has manifest many times his nonconformity with IVIC decisions," García Sucre told *Nature*. "In a certain sense he is an activist. In this situation it is not possible to be head of a lab — there must be a minimum of affinity with scientific politics." He adds that such personnel changes are routine, and that Mendoza still has all the rights of any IVIC staff member.

Mendoza says that he is unsure whether he will be dismissed entirely. "I don't think I will try to get reinstated as head. I am just basically trying to survive as a researcher," he says.

"I hope he will understand that the measure that has been taken is a mild one," says García Sucre, adding that in making fun of government officials, Mendoza has indirectly criticized president Hugo Chávez. Asked whether Mendoza will be fired outright, García Sucre says: "He should start to work in his lab instead of being in the newspapers all the time saying he is being victimized. Then I don't see any problem." ■

Michael Hopkin

See Editorial, page 702.

Federal agency rescinds primate stem-cell patents

Three key stem-cell patents have been revoked by the US Patent and Trademark Office. Scientists have complained that the patents, held by the Wisconsin Alumni Research Foundation (WARF), are so broad that they impeded US research.

The patents were granted between 1998 and 2006, but last July several nonprofit groups requested that they be re-examined, arguing that the work described in the patents was not sufficiently novel to deserve protection. The patents claim to cover primate embryonic stem cells — including human embryonic stem cells — derived by any means. The patent office agreed, on reconsideration, that the work was not novel enough to deserve a patent.

WARF has two months to contest the decision, and could then initiate a lengthy appeals process.

Nobel laureate named as NASA chief scientist

NASA's science directorate has been boosted by the appointment of astronomer John Mather as the agency's chief scientist.

Formerly a somewhat isolated position, the chief scientist will now have his own staff and work closely with the new associate administrator for space science, planetary scientist Alan Stern. Mather shared the 2006 Nobel Prize in Physics for work to probe the cosmic microwave background, the relic radiation left over from the Big Bang.

But the news is mixed for the agency. Last week, some members of Congress moved towards ousting NASA's inspector-general, Robert Cobb. His position is meant to oversee the agency from an independent standpoint, but several Congressmen charge that Cobb was too close to a former agency chief to perform his duties properly.

An independent government committee last week passed a report to the House Committee on Science and Technology alleging that Cobb had misused his authority and created a hostile work environment. Several lawmakers have asked President George W. Bush to fire Cobb and may hold Congressional hearings on the matter. Cobb has not responded to the allegations.

Arctic sea-ice cover reaches near-record low

This year, the extent of winter sea ice in the Arctic came close to breaking the record for being the smallest ever.

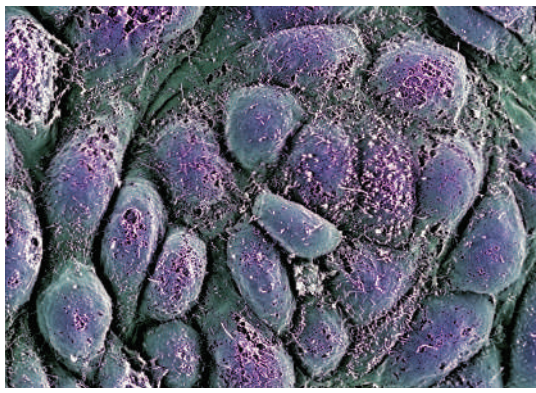
On 4 April, scientists at the National Snow and Ice Data Center in Boulder, Colorado, announced that the maximum sea-ice cover last month was 14.7 million square kilometres. The record was set in March 2006, when the Arctic was covered by just 14.5 million square kilometres. This finding is consistent with the marked trend of sea-ice shrinkage and thinning that has been observed in recent decades.

NASA, meanwhile, reported last week that less ice may be persisting from year to year in the Arctic, so less is available to replenish the perennial pack ice each winter. A study by Ron Kwok, of the Jet Propulsion Laboratory in Pasadena, California, found that only 4% of the seasonal ice that formed in the winter of 2004–05 survived the summer melt and helped build up the perennial ice in the winter of 2005–06.

UK report calls for change to stem-cell rules

British politicians have criticized their government for proposing a ban on the creation of 'chimaeric' embryos, made from human DNA inserted into an animal egg.

The Select Committee on Science and Technology, which is drawn from all the major parties, released a report on 5 April calling for a 'permissive' set of rules to be put in place that would allow the Human Fertilisation and Embryology Authority more freedom to adjudicate on the scientific merit of new embryological procedures.



The creation of human stem-cell lines from 'chimaeric' embryos is controversial.

M. STOJKOVIC/SPL

The committee hopes the government will change the rules when it releases a draft bill next month. Only that would end the limbo for researchers at King's College London and the University of Newcastle, both of which have applied for licences to create chimaeric embryos to generate human stem-cell lines.

University decides not to investigate smoking study

Officials at the University of California have decided not to examine a controversial 2003 research article disputing the dangers of second-hand smoke.

In recent months, questions had been raised about whether the article, published in the *British Medical Journal*, involved scientific misconduct because it relied on purportedly faulty smoke-exposure data (see *Nature* 446, 242; 2007).

Epidemiologist James Enstrom of the University of California, Los Angeles, (UCLA) was the lead author of the article (J. E. Enstrom and G. C. Kabat *Br. Med. J.* 326, 1057; 2003), which reported that spouses of smokers were no more likely to develop lung cancer or heart disease than were spouses of non-smokers. Enstrom denied any impropriety.

Late last month, UCLA officials reviewed concerns raised by the American Cancer Society about Enstrom's article, and decided that the allegations were not worth a formal inquiry or investigation.

Correction

In our News story "Letting the light in on Antarctic ecosystems" (*Nature* **446,** 9; 2007), the credit for the photo of the sea spider was inadvertently omitted. The picture should have been credited to P. J. López, University of Seville/Climant-Ecoantha.

Pictures of clouds reveal a Venus in twirls

Images beamed back from Venus are revealing complex cloud features above the planet that seem to vary from equator to pole.

In this picture, clouds near the Alpha Regio area — close to the planet's equator — show turbulent eddies, possibly caused by local convection set up by daytime heating from the Sun's rays. At higher latitudes, venusian clouds spread out into streaky bands, powered by the winds that zip around the planet in just four days.

The latest images were generated by a spectrometer on board the European Space Agency's Venus Express spacecraft, which has been probing the planet since it arrived there a year ago.

ESA/VIRTIS/INAF-IASF/OBS. DE PARIS-LESIA

Back to school

The US National Science Foundation may soon have to supply lab equipment to poverty-stricken high schools. **David Goldston** explains why some politicians want the agency to pay for Bunsen burners and test-tubes.

On 28 March, the House Committee on Science and Technology approved a measure that would put the National Science Foundation (NSF) in the business of equipping science laboratories in high schools that serve poor students.

Buying better high-school lab equipment is certainly a worthy goal and perhaps a legitimate use of federal money. Although education in the United States is mainly the province of state governments and local school districts, the federal government already provides billions of dollars to high schools that serve the poor via Department of Education programmes.

But paying for beakers is an unlikely task for the NSF. Supporting science and maths education at all levels is as much a part of the NSF's mission as supporting research, but its education efforts have traditionally focused on programmes to enhance teacher training, to create curricula and to promote experiments in educational approaches. The agency has not funded expensive purchases of books or equipment, largely because it tries to maximize the impact of its relatively small education budget. The NSF's annual spending on pre-college education can be counted up in different ways, but it is certainly less than $500 million a year — not much if it's to be spread around for routine purchases by every US school district.

So how did this legislation get on the agenda? The idea has been pushed over the past two or three years by the National Science Teachers Association and the American Chemical Society, which worked first with black and Hispanic members of Congress and then with members interested more generally in science education. The recent committee action was based on a bill sponsored by Representative Rubén Hinojosa (Democrat, Texas), who chairs the education task force of the Congressional Hispanic Caucus. One lobbyist told me: "The bill is a great way to get minority members interested in the National Science Foundation." And the bill moved quickly in the Democratic Congress because it was perceived as a way of showing concern for minorities.

The problem with this is not that the bill is targeted at poor and minority areas. The NSF runs several programmes to encourage minority students to study science, maths and engineering — and should probably run more, given the significance and extent of the shortfall. But

PARTY OF ONE

building interest in the NSF by presenting it as an agency that can dole out funding to local school districts, of any stripe, is asking for trouble.

Not only does the NSF lack the funding to take on such a role, it also lacks the staff to ensure the money would be used properly. The agency's inspector-general regularly complains that the agency's traditional grants in research and education are not followed up adequately. Saddling it with a hodge-podge of local efforts is likely to weaken the NSF's reputation for running targeted, prestigious and apolitical programmes — the very reputation that leads Congress to try to add to its responsibilities in the first place.

The school-lab bill is not the first time the NSF has been conscripted into minority politics. In 2003 and 2005, then-Senator George Allen (Republican, Virginia) offered a bill to have the NSF buy computers for colleges that were created to serve, or which now serve, predominantly minority students. Allen gave the task to the NSF because of the agency's reputation and because he served on the committee that oversees the NSF. The Senate passed the bill twice, but it always stalled in the House. (As chief of staff of the science committee, I was involved in deciding how to handle both the Hinojosa and Allen bills in past Congresses.)

When Allen's re-election effort ran into trouble last year because of charges of racism, the Republican leadership made some efforts to see whether the bill could be revived at the last minute. Nothing came of it, largely because the House Committee on Education and the Workforce remained opposed.

Proponents of the Allen bill tried to justify their efforts by citing a study, by the National Association for Equal Opportunity in Higher Education, that they said showed the need for computer equipment at minority institutions. But the study's main point was that the institutions did not know how to make the best use of computer equipment they already had.

Similarly, advocates of the Hinojosa bill cite the National Academy of Sciences' 2005 publication *America's Lab Report: Investigations in High School Science*. That report does point out that non-Asian minorities generally have access to poorer labs and get less time in them, but its main conclusions do not focus on equipment. Instead it notes how limited education research, inadequate teacher training, bad curricula and misguided science standards result in labs being used in ways that are unlikely to enhance education. Equipment is the least of it.

But Congress is best at providing tangible goods, so that tends to be the focus of legislation, regardless of whether it is the greatest need.

Staff members on the science committee were sensitive to some of these issues, so the approved version of the Hinojosa bill is an improvement over the initial legislation. The programme is now authorized as a $5-million 'pilot' rather than the $50-million-a-year ongoing programme originally proposed. And applicants must couple the equipment purchases with teacher training. The bill requires anyone who wins federal funding to match it, and the committee may encourage, but not require, that the non-federal portion pays for the actual equipment purchases.

This approved version was put forward by Representatives Eddie Bernice Johnson (Democrat, Texas), who belongs to the Congressional Black Caucus and has long been a supporter of the NSF, and Vernon Ehlers (Republican, Michigan), a physicist with a longstanding interest in science education. They offered it as an amendment to a larger science-education bill that the House is likely to take up later this spring.

That bill, in turn, is likely to become part of a larger negotiation with the Senate over legislation meant to keep America competitive in the world economy. Congress will then have to decide whether a measure that assigns the NSF the task of buying test-tubes for the nation's most troubled high schools is the best way to have the agency contribute to US competitiveness. ■

David Goldston is a visiting lecturer at Princeton University's Woodrow Wilson School of Public and International Affairs.

BUSINESS

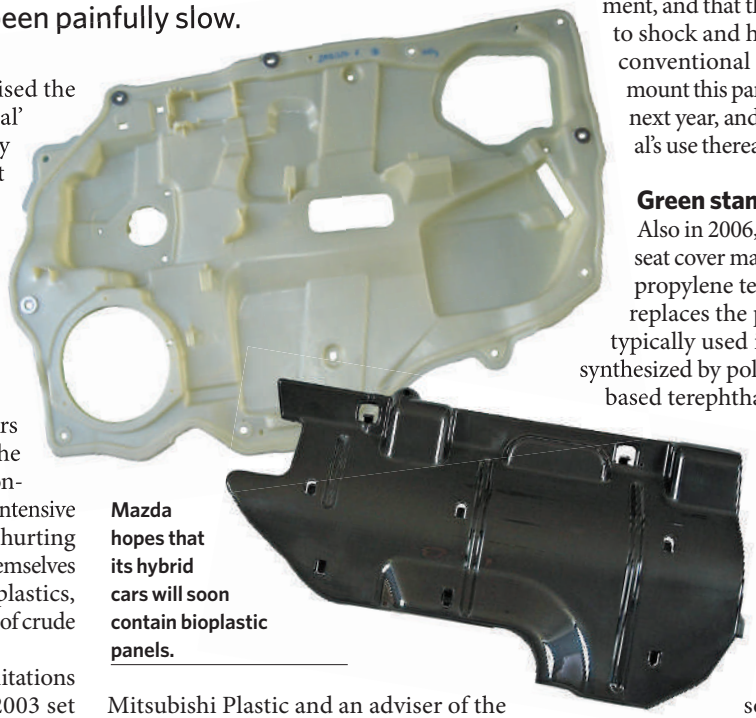
Plastic promises

Car companies know that plastic parts made from plants will appeal to 'green' customers. But as **Ichiko Fuyuno** reports, their progress has been painfully slow.

Back in 1941, Henry Ford surprised the public by unveiling a 'biological' car. The legendary car-company boss, who had a long-standing interest in environmentally sound materials, unveiled a prototype of plastic panels made from soya beans, wheat and hemp attached to a tubular frame at a motor show in Dearborn, Michigan. But after the United States joined the Second World War, the project slipped quietly from view.

Sixty-six years on, Japan's car makers are picking up where Ford left off. The hydrocarbon-based plastics conventionally used in the car industry are energy-intensive to produce and difficult to recycle — hurting companies' strenuous efforts to paint themselves green. That increases the lure of bioplastics, which are produced from plants instead of crude oil and generally biodegrade over time.

But these materials have severe limitations — so severe that Toyota, which in 2003 set ambitious goals for the commercial use of bioplastics, is considering scaling back its commitment. If it does, all it will have to show for years of bioplastics research and development will be a floor mat and a cover for a spare tyre.



MAZDA

Mazda hopes that its hybrid cars will soon contain bioplastic panels.

Mix it up

Typically, bioplastics are made by fermenting glucose obtained from corn, sugarcane or other plants into lactic acid, which is then polymerized to make the polylactic acid that sets to form the plastic. But the materials tend to have poor mechanical properties and degrade quickly, so manufacturers often mix in chemical additives or fibres — or blend them with conventional materials to improve their strength and durability.

The market for bioplastics remains modest. According to Japan's Biodegradable Plastics Society, they account for less than 1% of the nation's consumption of plastics, with almost all of that going to low-grade applications such as food packaging.

Still, Japanese companies are keen to use bioplastics to make things such as computer cases and mobile phones, as well as in the car industry — a sector of bulk plastics users that has high performance specifications.

"It is not easy," concedes Isao Inomata, of Mitsubishi Plastic and an adviser of the society, "but technological advances have made bioplastics available for higher-end products than food packaging, and car makers are actively working on them."

Toyota first started to develop bioplastics in 1999, and four years later it started to sell accessories made from these materials. They designed a floor mat made from 100% polylactic acid, and a spare-tyre cover made from polylactic acid and fibres from kenaf, a plant often used in paper.

Great expectations

In a 2003 Environmental and Social Report, a division of Toyota said that it hoped to produce two-thirds of all of the world's bioplastics and earn ¥5 trillion (US$40 billion) in sales per year by 2020, but the company says this was never its official business plan. And in 2006, Toyota withdrew the floor mat from production. It is now considering a reorganization of its bioplastics business, according to spokeswoman Atsuko Watanabe.

Toyota's rival Mazda has made the most progress in incorporating bioplastics into vehicles. In 2006, it released a prototype of an instrument panel — made from 88% corn and 12% petroleum — for one of its hybrid cars. The company says it used an additive that allows the plastic to solidify into small, quick-forming crystals — slow crystallization is a problem with many bioplastics because it increases the costs of making parts from them. The company also says that the panels can be manufactured with conventional injection-moulding equipment, and that they offer as much resistance to shock and heat as products made with conventional materials. Mazda plans to mount this panel on its production hybrids next year, and hopes to widen the material's use thereafter.

Green standards

Also in 2006, Honda released a prototype seat cover made of a biofabric called polypropylene terephthalate. The bioplastic replaces the polyethylene terephthalate typically used in car seat covers, and was synthesized by polymerizing 60% petroleum-based terephthalic acid with 40% of a glycol derived from corn. The company says that its material will reduce carbon dioxide emissions on disposal by 5 kg per vehicle, or half, from those of its existing plastic seat covers. It has also devised a way to weave the cloth that gives the material a soft texture while retaining its strength and lightness. "We hope users are going to choose the biofabric because it feels more comfortable," says Keiichi Araki from the design development division at Honda. The company plans to use the biofabric as standard equipment on fuel-cell cars from around 2008.

Difficulties with the strength and durability of plant-based plastics have caused the Biodegradable Plastics Society to effectively 'lower the bar' and allow its members to produce what it calls 'biomass plastics' that contain as little as 25% of plant-based raw materials.

That has led sceptics to cry foul. Yoshio Watanabe, a motor-industry analyst at Mizuho Securities in Tokyo, doubts that companies are genuinely committed to the use of plant-based plastics. "I have never heard of concrete long-term business plans about bioplastics from any car-makers," he says.

But Takahiro Tochioka, an engineer at Mazda who specializes in bioplastics, insists that real progress is being made. For instance, he says that his team has started to develop enzymes that can control the time that bioplastics take to degrade. "About a year ago, many people felt that using bioplastics in cars was unrealistic. But what we have done in the past year has brought back a positive prospective," he says. ■

T. BOLSTAD/PANOS

How to survive a warming world

African communities have been adapting to climate change for millennia. **Jim Giles** reports on the strategies that seem most effective.

Four decades ago, drought arrived in El Fasher, an impoverished state capital in western Sudan. Rural communities there had relied on weak rains to raise crops in sandy soils. But rainfall has been below average ever since, displacing a million people in the area and forever altering the lives of many more.

Yet the people of El Fasher managed to adapt and survive. They built low earth embankments, known as *trus*, around their villages to hold water and irrigate crops of sorghum and vegetables. They developed new planting methods, digging through the layers of sand to the fertile ground beneath. And they introduced a wider range of crops, from citrus fruit to tobacco, both to broaden their food sources and to sell.

Today, El Fasher is facing a vastly different challenge, as it is at the heart of the bloody Darfur conflict. But elsewhere in the world, El Fasher's experience during drought should prove useful. Humans have adapted to changing environments for millennia, and many communities contain a wealth of knowledge about how to beat climatic odds and survive when rains fail or floods sweep away crops. So far, there have been few systematic studies of these survival strategies; but by tapping into communal knowledge, researchers are beginning to tease apart the actions that determine whether or not a community will make it in the face of change.

The process is formally known as 'adaptation', a term that is becoming a political buzzword even among environmentalists, who once saw it as a distraction from the business of cutting carbon dioxide emissions. Now, with climate change expected to have unprecedented effects on people worldwide (see page 706), adaptation can no longer be shunned (R. Pielke *et al. Nature* **445,** 597–598; 2007). "People talk about adaptation as if it's a new invention," says Guy Jobbins, a senior programme officer at Climate Change Adaptation in Africa, a research organization based in Cairo. "It's not. People in Africa have been adapting for thousands of years."

It's not only Africa that will have to adapt, but the continent is often singled out because many of its residents live in precarious circumstances. In some countries, poor soils and a lack of government support already make many vulnerable to drought or extreme weather, so they have little to fall back on should climate change make things worse. Yet that is exactly what climate models predict is about to happen. "Drought and other climate disturbances exact an unacceptably high and reducible toll on the people of Africa, a toll that is likely to grow with climate change," says Balgis Osman-Elasha, a climate-policy expert at the Higher Council for Environment and Natural Resources in Khartoum, Sudan.

Africa's climate is poorly understood, in part because of patchy historical weather data. But the results that are available give cause for concern. The Sahara, for example, is expected to warm at rates greater than the global average of about 0.2 °C per decade during the early part of this century. Computer models also predict that eastern Africa will get wetter and see an increase in extreme rainfall events, and that

parts of southern Africa will become drier.

Elsewhere on the continent, things are less certain. Richard Washington, an expert in African climate at the University of Oxford, UK, says that predictions are vague for El Fasher and the rest of the Sahel — the east–west band of semi-arid land between the Sahara and the greener landscape of central Africa.

This is partly because the thunderstorms of the monsoon are difficult to simulate in climate models. Capturing the effect of dust blown from the deserts of North Africa is another challenge the models struggle with. Consequently, says Washington, "you can get any result you like" with a climate simulation. If all models are averaged, a wetter future is predicted; if the selection is limited to those that closely simulate the current climate, the future looks drier than today.

Changeable weather

Data, and hence predictions, are better in South Africa, where some regions are already experiencing changes similar to those predicted by some of the better models. A team led by David Thomas, of the University of Oxford, has been studying the village of eMcitsheni, in the eastern part of the country, where rains have become more uncertain during the past 50 years. More heavy rainfalls have been arriving early in the wet season, before Christmas, and fewer at the tail end, around March.

Researchers can't say whether such changes are due to natural variability or rising carbon dioxide levels, but that matters little to the communities experiencing new weather patterns. Nor does it matter to those interested in understanding adaptation. As similar climatic changes are expected to occur in the future, villages in the region form a kind of natural laboratory in which to study the ongoing effects of climate change.

Several adaptation strategies are used at eMcitsheni, a rural area accessible only by dirt road that is home to about 300 households who depend on agriculture and livestock. Crops are planted farther apart so that more moisture is available for each row, increasing the likelihood that they will survive a period of drought. Corn (maize) varieties that mature faster have been brought in, again limiting the threat of dry spells. Local people have also set up a commercial cooperative: if the village produces excess maize in a certain year, the group works together to transport the crop to markets, earning money that can be used to buy food when yields are lower.

D. ROSE/PANOS

Different irrigation strategies have helped people in Africa's Sahel region survive persistent drought.

But would such strategies work in other places? Many development experts are cautious about extrapolating from small studies to broader rules about how communities in general should adapt. Climate is far from the only factor changing people's lives, and given the complexity associated with even simple changes, it is dangerous to think about a magic-bullet solution to adaptation, says Jobbins.

> **"In general, communities seem to have adapted best when working as a collective rather than as individuals."**

Yet some general principles are starting to emerge. To that end, the Oxford team has studied three other rural communities in South Africa and Mozambique. All of these villages rely on crops and livestock, and all have seen substantial climate change during the past decade. In the Limpopo province, near the South African border with Zimbabwe, the wet season has started later over the past 50 years and included more dry spells. Further west, in the region near the town of Mafikeng, farmers have seen more rainy days early in the wet season, but more overall variability in rainfall between successive seasons.

In general, communities seem to have adapted best when working as a collective rather than as individuals. Overall, the team found four main strategies of adaptation: changes to agricultural practices; the formation of social networks; commercial projects, such as investing in livestock; and seeking work in distant areas. The first three of these strategies rely on people working together to better their community.

In eMcitsheni, for example, people developed communal horticultural projects. Local women created jointly run irrigated gardens that, because the risk was shared among the collective, allowed them to diversify into crops they might not have attempted to grow on their own, such as potatoes. When heavy rain damaged traditional corn plots, these gardens could help to compensate for what was lost. Surplus crops were sold and the proceeds invested in pumps.

Such sharing seems to work for many groups, says Henny Osbahr, a geographer at the University of Oxford. "It's complex," she says, "but we did see generic characteristics: strong informal institutions and networks of reciprocity."

Of course, many groups' ability to adapt is limited by factors beyond their control, such as lack of skills or money. Osman-Elasha has studied communities in El Fasher and other parts of Sudan, and says that various 'essential resources', such as access to spare machine parts, were cited repeatedly as problems by local people.

Joining forces

When communities work together they are better able to interact with outside organizations such as government agricultural officials and donor agencies. And several larger development agencies are making adaptation a formal part of what they do. The US Agency for International Development (USAID) in Washington DC, for example, has begun funding the introduction of rice varieties in southern Mali that are better suited to shorter rainy seasons, because farmers there have reported experiencing hotter and drier conditions in recent decades. USAID will release an adaptation handbook for all its programme officers this May; the World Bank is also developing adaptation information for its staff.

The agencies, along with local and national governments, face an enormous struggle. With wars, poverty and the AIDS epidemic, Africa is perhaps the region least well equipped to cope with climatic disturbances. The focus on building on local knowledge doesn't offer all the answers, but to adaptation experts it at least offers a way forwards.

"People felt that too much time spent on adaptation was taking attention away from doing something about the problem," says Emma Archer, a climate researcher at the University of the Witwatersrand in Johannesburg, South Africa. "That it was letting people off the hook. But now we know that we're committed to change. We're at an exciting stage." ■

Jim Giles is a senior reporter with *Nature*.

A. COOPER/ALAMY

A world melting from the top down

Despite years of speculation, little can be said for sure about the future of the Arctic's permafrost. But that's no grounds for complacency, reports **Gabrielle Walker**.

Some of Phil Camill's trees are drunk. Once, the black spruce trees on the plots of woodland that he monitors in northern Manitoba stood as straight and honest as pilgrims. Now an ever-increasing number of them loll about leaning like lager louts. The decline is not in the moral standards of Canadian vegetation, but in the shifting ground beneath their roots. Once it was all hard, solid permafrost. Now much of it has thawed into a soggy sponge that no longer provides a steady footing for the trees. Some contrive to grow at screwball angles; others have drowned and been replaced by floating mats of mosses or sedges. "It is really easy to tell when the permafrost has gone," says Camill. "The vegetation changes right before your eyes."

Camill, an ecologist from Carleton College in Northfield, Minnesota, has used those changes to trace the rate at which the permafrost is disappearing. In their desperate attempts to buttress themselves upright, his leaning spruces put on extra wood on the downslope side of their trunks. Counting the asymmetrical tree rings that result and measuring the distance of each tree from the current boundary of the permafrost gives a measure of the pace of change. The results are shocking. An average warming across his sites of 1.3 °C since 1970 has brought with it a trebling of the thaw rate. In some places the permafrost's perimeter is retreating by 30 centimetres a year[1]. If this trend continues, Camill estimates that no permafrost will be left in any of his five sites by the end of the century.

Thawed-out permafrost has already undermined buildings, highways and other infrastructure from Alaska to Siberia. The damage is one of the most visible effects of warming temperatures on human activities. But the effects on natural systems are to some extent more worrying. Buildings can be rebuilt, asphalt relaid and agricultural practices changed through adaptation, given the right policies and priorities (see page 716). But changes in the vegetation and, crucially, in the soils of the frozen northern landscapes might not be so easy to cope with. The soils of the Arctic are crammed with organic matter — a frozen reservoir of beautifully preserved roots, leaves and other raw material that may contain as much carbon as the whole atmosphere. They are quite unlike soils from more temperate regions, which are mostly made up of the parts that the bacteria cannot digest. "We are unplugging the refrigerator in the far north," says Camill. "Everything that is preserved there is going to start to rot."

For decades environmentalists have worried about the possibility of this great putrefaction. It has become perhaps the most cited example of the biogeochemical feedback that could drastically worsen the effects of anthropogenic climate change. The idea is that humans increase levels of carbon dioxide in the atmosphere, warming the permafrost, which in turn releases yet more carbon, warming the world — and the permafrost — further still in an ever-escalating positive-feedback loop.

However, although such feedback has been discussed for almost as long as the threat of global warming has been taken seriously by scientists, the lack of firm data on the subject is striking. "There is a lot that we don't know at this point," says Walter Oechel from San Diego State University in California. "People haven't quite pulled the whole picture together yet — but what we do know is that the potential amounts are huge and very, very scary."

The big picture

There is no doubt that the Arctic is heating up. Vladimir Romanovsky from the University of Alaska Fairbanks has collated borehole- and air-temperature data from throughout the Arctic[2]. He found that only one region in the Arctic had not warmed over the past 30 years — and

in the 1990s even that region joined the trend. Some places are warming at more than twice the global average rate. Romanovsky recently received a US$1 million grant to take this monitoring work further with a network of stations in North America and Russia.

The effect that this warming will have on the permafrost and its stored carbon will vary from region to region. Not surprisingly, the most dramatic signs of thaw have come from the fringing, southerly regions of discontinuous permafrost — such as Camill's research sites — where the frozen layer is only a few metres thick and average temperatures are already within a whisker of the melting point. In the colder Arctic the permafrost can be hundreds of metres thick and it is harder to know what to expect. In principle, the thawing might be quite slow, with the warmth at the surface being transmitted gradually into the colder depths. In practice, things are probably more complex, and in places more precipitous.

Patchy progress

Permafrost is defined as ground in which the temperature is less than 0 °C for at least two successive years. But the ground in question does not have to be at the surface. In most places the top part of the soil thaws during the summer, providing plants and microbes with an 'active layer' above the permafrost in which they can flourish and decompose the defrosted organic matter. The probability that the active layer will deepen — putting a larger stash of carbon up for grabs — or that the permafrost will thaw completely depends on the type of vegetation and soil. Thus, thawing can accelerate rapidly if a fire passes through a dried-out forest in the uplands, or if the soil contains enough ice that thawing causes it to collapse, creating a crater-scarred 'thermokarst' landscape.

> **"We are unplugging the refrigerator in the far north. Everything that is preserved there is going to start to rot."**
> — Phil Camill

Researchers expect to see the first signs of thaw in a deepening of the active layer. But complications from terrain, vegetation and other local conditions mean that data need to be collected continuously over several decades to pick up such a trend. Unfortunately, the relevant measurements so far have been patchy and sporadic. In 1998, various organizations involved in permafrost research banded together to receive funding as the Circumpolar Active Layer Monitoring (CALM) programme, hosted at the University of Delaware in Newark, to tackle the problem. The network now includes 125 active sites and has participants from 15 countries. Still, it is unlikely to bear fruit in the form of spotting unequivocal trends for some time.

If measuring the damage done so far is hard, predicting its future course is even harder. In 2005, two researchers from Boulder, Colorado — David Lawrence from the National Center for Atmospheric Research and Andrew Slater from the Cooperative Institute for Research in Environmental Sciences — published the results of the first attempt to project the fate of the permafrost through the twenty-first century using the climate predictions of a general circulation model (GCM). Their results made dramatic headlines. Using figures from one of the 'high emissions' scenarios developed by the Intergovernmental Panel on Climate Change (IPCC) in the model resulted in 90% of the northern permafrost disappearing by 2100. Of 10.5 million square kilometres of permafrost around today, only about 1 million made it through the century. Even more worryingly, running the model with a 'low emissions' scenario still wiped out 60% of the permafrost, suggesting that severe losses are inevitable no matter which policies are followed. And the model did not take into account any further warming from carbon given off in the thaw[3].

P. REKACEWICZ, UNEP/GRID-ARENDAL; INT. PERMAFROST ASSOC., 1998. CIRCUMPOLAR ACTIVE-LAYER PERMAFROST SYST. (CAPS), V.1.0.

Thick and thin

The model has since attracted some criticism, most notably because its permafrost is a mere 3.4 metres thick throughout the Arctic, which is far from the hundreds of metres present in some regions. However, Lawrence says that he has since re-run the model for a mid-way emission scenario and dealt with some of the other criticisms at the same time, and the results remained more or less the same. He says that the model captures many important aspects of the Arctic system — including the hydrology and physical properties of the soil — and that the point is not so much the actual percentage of loss but the overall principle that a frighteningly large amount of permafrost could be vulnerable to quite small changes in climate. "In this field you have to accept that we won't have a perfect knowledge of what's happening up there," he says. "But we should be able to capture the fundamental properties. And so far the model shows that a major change is going to happen to the Arctic."

Perhaps the most intriguing of the complications that Lawrence's original work did not address is the suspicion in the minds of some researchers that the permafrost itself is putting up a defence against the thaw — a set of negative feedbacks. For instance, Oechel points out that warmer temperatures lead to a thicker layer of moss on the surface of his research sites in the Alaskan tundra. Because

M. HOSHINO/MINDEN PICTURES/FLPA

Mosses act as insulators for underlying permafrost.

moss is a superb insulator, especially when dried out by surface warming, the thickened vegetation helps to shield the frozen soil beneath from the warmth above. In one of the north–south transects he studies on Alaska's North Slope, the warmest, southernmost section has both the thickest layer of moss on the surface and the shallowest active layer.

Frederick Nelson at the University of Delaware points to another self-preserving feature of permafrost. The base of the active layer, he says, can become especially icy because water draining down there will pool above the impermeable layers below. The richer this layer is in ice, the more difficult it is to thaw, preventing the active layer from deepening further[4]. Even snow cover does its part. Like moss, snow is an effective insulator. But because it falls in the winter, it works in the opposite direction, shielding the soil from cooling further in air temperatures that can be as low as –40 °C. Camill noticed that the snow at some of his sites has thinned in the past few decades, which may cause the soil to grow colder in winter than it used to and thus store up protection against the heat of the following summer. Increased forest cover can have a similar effect, causing snow to be thinner beneath the trees than it would have been on open ground. "At face value you would expect the permafrost to start thawing really rapidly as temperature rises, but these feedbacks can keep it around longer than you would expect," says Camill.

Pushing the boundaries

But these effects can't last forever — and although they might forestall thawing, the sudden change in conditions when one or more of them fails might lead to quicker thawing thereafter. A few really hot summers could break through the ice barrier. And Oechel is already worried about his mosses. He has noticed that they are highly sensitive to direct sunlight, and now that the Arctic has fewer cloudy days the mosses could well begin to suffer. Loss of any or all of these protections would allow any thaw to accelerate. "It is difficult to push permafrost over the threshold of thawing," says Romanovsky. "But after tipping it will go by itself".

Uncertainties about thawing obviously complicate the question of how much greenhouse gas the permafrost will emit — but they are not the only complications. Oechel has been tracking the carbon balance of his Alaskan tundra for several decades. After a serious bout of warming in the late 1970s, he saw the region emit a mighty pulse of carbon dioxide — a pulse that is now showing signs of tailing off. He suggests that one possible reason for the decline is that the initial orgy of decomposition spurred by the warming released nitrogen-containing nutrients into the soil. The plants have now responded to the additional nutrients by growing more and taking up carbon dioxide in the process. Oechel thinks that in the summer some of his sites now take up more carbon than they emit, although when winter emissions are added in they are still sources, not sinks, overall[5]. Another factor is the northern march of the treeline. As the region warms, the growth of new forests over what was once tundra could also help to reduce the net release of carbon.

However, even if the surface ecosystem starts to take up more carbon, the old, dead material frozen in the soil still needs to be considered. What's more, the overall amount of carbon emitted is not necessarily the whole story. Torben Christensen at Lund University in Sweden also has data on the carbon balance that span several decades, in his case from a low-lying mire in the patchy permafrost of northern Sweden[6]. As the permafrost has steadily thawed, the ground has grown soggier; the previous hummock vegetation has been replaced by sedges, which are better at tolerating wet roots. The water seals the underlying soil from the air, which means that decomposition has to proceed without the benefit of oxygen. This slows things down, and has reduced the region's carbon emissions by 13% since 1970. However, the carbon that is released by this oxygen-free decomposition comes out in the form of methane, rather than carbon dioxide, and methane packs a far greater warming punch than its oxidized sibling. So the overall greenhouse effect of the mire has actually gone up by a disturbing 47%.

V. ROMANOVSKY, UNIV. ALASKA FAIRBANKS

Thawing ice under the surface can cause the landscape to collapse.

Back to biology

These are the sorts of issues that future attempts to model the process will need to take into account. Lawrence says that the next phase of this work will be to treat the biology more carefully, incorporating carbon and nitrogen cycles into the models, allowing the vegetation to respond to the changes in climate, and modelling sources of methane. According to Nelson at least half a dozen groups around the world are planning to develop their GCMs to address at least some of these aspects of the permafrost thaw, and the plan is to have a much more extensive set of predictions in the next IPCC report, in five years time.

Knowing more about the thaw, though, will allow useful predictions of carbon emissions only if researchers can also quantify the amount of organic matter in the soils. A series of workshops under the auspices of the Global Carbon Project and the International Polar Year is aiming to answer those questions. Current estimates of the amount of carbon that might be in play range from 350 gigatonnes to more than 900 Gt;

S. ZIMOV

Fire and ice: decomposition under thaw lakes is releasing large amounts of methane, which bubbles to the surface and adds to carbon emissions.

by way of comparison, the atmosphere contains 750 Gt or so. The estimates at the high end of the scale are based on the discovery of a new, vast pool of buried carbon — a type of wind-blown soil called yedoma, which was laid down over large tracts of northern Siberia in the ice ages.

Rich sources

Stuart Chapin, also from the University of Alaska Fairbanks, says that the yedoma soil is extraordinarily rich in carbon. Last year, he and his colleagues estimated the range and possible thickness of this layer and calculated that it alone could contain 450 Gt of carbon, compared to the estimated 350–450 Gt in the rest of the Arctic[7]. This number, says Chapin, is probably "only good to within a factor of two"; but even half of such a huge amount would be significant, whereas twice as much hardly bears thinking about. Camill points out that humans release around 9 Gt of carbon per year from fossil fuels and deforestation. "If just 1% of [the possible 900 Gt in the yedoma] is decomposed in a warmer world it would be as if we doubled our current rate of emissions. That's what is alarming."

Ominously, the first signs that parts of the continuous permafrost might now be thawing have come from lakes that overlie this carbon-rich yedoma. Katey Walter from the University of Alaska Fairbanks and colleagues have been tracking the methane that bubbles out of thaw lakes in northern Siberia. These lakes don't necessarily arise from global warming. Any local disturbance can trigger a temporary thaw in the permafrost. As the ice melts, the ground sinks and fills with water. The lakes then tend to migrate across the landscape, eroding away their margins, and can last as long as a thousand years; their sideways motion allows them to eat through permafrost much more quickly than would a steady heating pulse heading down from the surface. And decomposition in their oxygen-free depths and the thawed sediment beneath will produce methane, not carbon dioxide.

> **"It is difficult to push permafrost over the threshold of thawing, but after tipping it will go by itself."**
> — Vladimir Romanovsky

How much methane nobody realized until recently — mainly because it bubbles out from random parts of the lake and disappears unnoticed into the air. Walter and her colleagues managed to catch this methane in the act, by noting where the bubbles emerged when the lake froze over in the winter, and then leaving instruments to catch the emissions throughout the following year[8]. They calculate that bubbling lakes from northern Siberia are already responsible for nearly four million tonnes of methane a year, and that the amount is on the rise. Warmer temperatures mean that the lake area in Walter's study region has increased greatly in the past few decades, leading to a rise in methane emissions of nearly 60%. And that methane seems to be coming from the depths of the permafrost: its lack of carbon-14, an isotope continually made in the atmosphere that takes thousands of years to decay, suggests that the organic matter beneath the lakes has been stored away for a very long time.

This and the Arctic's other warning signs make it increasingly urgent that researchers resolve their remaining questions about the fate of the permafrost. And those answers won't come a moment too soon. "We have been asleep at the switch," says Oechel. "If you look at the things that were said in the 1970s about the Arctic's response to increasing CO_2, the place we were off is not that we overstated or were overly pessimistic, but that we were not aggressive enough about the predictions. To me, the precariousness of the situation is now clear. We are in a world of hurt." ■

Gabrielle Walker is the author of *An Ocean of Air*.

1. Camill, P. *Clim. Change* **68,** 135–152 (2005).
2. Romanovsky, V., Burgess, M., Smith, S., Yoshikawa, K. & Brown, J. *EOS Trans. AGU* **83,** 589–594 (2002).
3. Lawrence, D. M. & Slater, A. G. *Geophys. Res. Lett.* **32,** L24401 (2005).
4. Shur, Y., Hinkel K. M. & Nelson, F. E. *Permafrost Periglac. Process* **16,** 5–17 (2005).
5. Oechel, W. C. *et al. Nature* **406,** 978–981 (2000).
6. Johansson, T. *et al. Glob. Change Biol.* **12,** 2352–2369 (2006).
7. Zimov, S. A. *et al. Geophys. Res. Lett.* **33,** L20502 (2006).
8. Walter, K. M., Zimov, S. A., Chanton, J. P., Verbyla, D. & Chapin, F. S. *Nature* **443,** 71–75 (2006).

AL'S ARMY

Members of the public are taking to the streets to spread Al Gore's message of climate crisis. **Amanda Haag** meets the foot soldiers of global warming.

It's after hours at Monarch High School in Louisville, Colorado, and the hallways have fallen silent. In one classroom, a handful of high-school science teachers sit at desks usually occupied by biology students. Mark McCaffrey is giving a talk on how to teach global warming, and he points to a PowerPoint projection of the full-Earth shot snapped by Apollo 17 astronauts: the iconic 'blue marble'. "Everything that's ever happened in human history has happened on this fragile little spaceship Earth," he says, almost reverentially.

If this seems familiar, it could be because it sounds like Al Gore. McCaffrey is one of roughly 1,000 volunteers who, since last September, have been through a two-and-a-half-day training session with former US vice-president Gore and his staff in his home town of Nashville, Tennessee. The point is to spread the message of Gore's Oscar-winning documentary on climate change, *An Inconvenient Truth*. Trainees learn how to manoeuvre through the science of Gore's 300-plus slide presentation and to discuss weighty topics such as rising carbon-dioxide concentrations, the physics of hurricane intensity and the mechanics of sea-ice retreat.

They then return to their communities hoping to reach as many others as possible, in high schools, churches, city council meetings, businesses and retirement homes. McCaffrey, whose day job involves science education and outreach for a joint institute at the University of Colorado, Boulder, says that one of his motivations was to improve his understanding of climate science so that he could respond to naysayers. "The sceptics would derail me and I'd get flustered and not know how to respond," he says. "Sometimes they're scientists from outside the field of climate, but they know enough about how to throw up uncertainty and plant doubts in people's minds."

Gore calls the trainees his "cavalry", but a more apt name might be missionaries, given the fervour with which they approach their roles. One volunteer, from Hackett, Arkansas, signs off his e-mails as "Robert McAfee, Climate Change Messenger". Gary Dunham, an independent voter from Sugar Land, Texas, says he had a near-religious conversion while watching *An Inconvenient Truth*. "I went to see the movie intrigued by what it was about but certainly not believing in the global-warming message," he recalls. "Within 15 minutes I completely changed my viewpoint. I don't think I've heard a political speech that really motivated me to get up and do something since John Kennedy's day." And volunteer Reggie Allen of Keller, Texas, says he sees the need to disseminate the "truth" about global warming as a mission akin to the civil-rights movement, for which his parents used to march after church on Sundays.

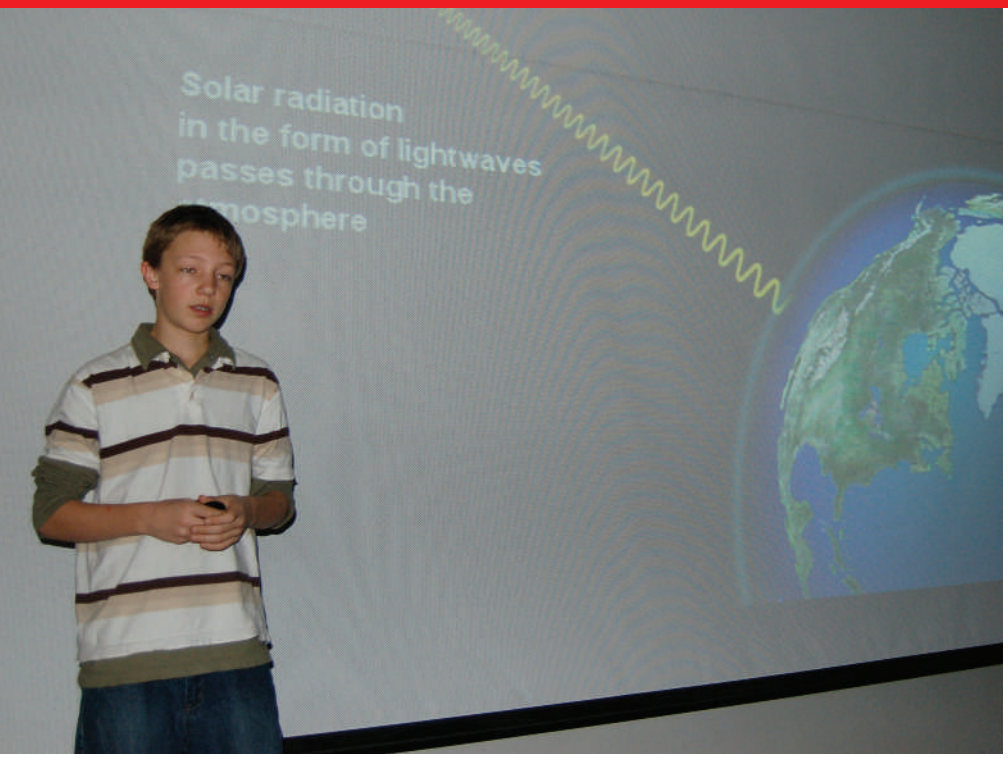
PARAMOUNT CLASSICS

Al Gore's Oscar-winner has made climate change a hot topic for debate in the United States.

The volunteers were chosen from several thousand applicants, and include a middle-school student, priests, mayors, nuclear engineers, right-wing conservatives, Wal-Mart employees, Miss Rhode Island and Cameron Diaz. No matter what their walk of life, their motivation is the same: to tell their friends, families and neighbours that human activities are altering global climate and that each person can do something about it.

The mission begins in Nashville where, on the first full day, Gore himself leads 90% of the training, walking volunteers through the science slide by slide. He takes questions, and a scientist is always present to help answer them. Gore's team includes a rotation of four scientists, including Michael MacCracken, chief scientist for climate change at the Climate Institute in Washington DC and a longtime contributor to the Intergovernmental Panel on Climate Change, and glaciologist Richard Alley of Pennsylvania State University.

On the second day, the trainees break into smaller groups to practise giving the presentations themselves.



A. L. HAAG

14-year-old Alex Budd is the youngest volunteer spreading Al Gore's message on climate change.

"They blow you out of the water because they're speaking from their hearts," says Carey Stanton, senior director for education at the National Wildlife Federation, and one of Gore's staff. "When they're backed up by really knowing the science, they're very good."

Trainees are encouraged to tailor their talks to individual audiences while still preserving the framework of Gore's presentation. For instance, the inspirational images of Earth from space are expected to bookend the presentations, as they do the film. In his presentation, McCaffrey shows about a third of the available slides.

Made to measure

McCaffrey veers away from the original presentation mainly to use examples he sees as effective for spreading the message to young people. "If you just have a laundry list it's going to go in one ear and out the other. At least that's my experience with a teenager in the house," he tells his audience. He also mentions opportunities for students to take the lead in being part of the solution, such as by initiating carbon-neutral school programmes.

The Gore campaign has garnered a near cult-like following. In December, the left-wing advocacy group MoveOn.org campaigned to have the documentary aired on the same day and time across the United States. Training sessions have taken place in Australia and the United Kingdom. And last week, Gore supporters flooded the website of Step it Up, a group planning a 'National Day of Action on Climate Change' on 14 April, and nearly shut the site down.

But the movement hasn't had a warm reception in all quarters. In November, the producers of the film sought to have it mass-distributed to high-school classrooms, offering to deliver 50,000 free DVDs to the National Science Teachers Association. The association baulked, saying that it doesn't send out unsolicited material to its members and that doing so would "constitute an endorsement" that might trigger other "special interests" to ask it to distribute material. Laurie David, one of the film's producers, publicly criticized the association's unwillingness to distribute the DVDs — even though the group offered to post links to the film on its website and to make it available to anyone who requested it.

And in January, the school board of Federal Way, a district near Seattle, Washington, made national news when it placed a temporary moratorium on showing *An Inconvenient Truth*. A parent had complained that the film presents only one side of the global-warming debate, and district policy states that teachers who choose to show material containing "bias" must also present a "credible, legitimate opposing view" and that the principal and superintendent must grant permission. The moratorium has since been lifted.

A. L. HAAG

"Educators have a particularly important job in communicating the basics of climate."
— Mark McCaffrey

Even in school districts as left-leaning as Boulder, a quieter undercurrent of dissent brews about how — and how much — global warming should be taught in the classroom. Some Boulder Valley teachers showed *An Inconvenient Truth* in class and met with resistance from parents and other teachers afterwards. "Teachers tend to shy away because we don't have the political support and backing for controversial issues," says Kristin Donley, who coordinates science curricula for the Boulder Valley School District. She is working to develop a unit on climate change for her classes. Some of the basic concepts, such as the carbon cycle and the greenhouse effect, are taught by default in physical sciences, Donley says. But the curricula offer a lot of leeway, and teachers can choose whether or not to broach the topic of global warming.

Lesson plan

McCaffrey came to the school after hearing that teachers in the district were interested in including climate change in their lessons. As he flips through his final slides, he charges the teachers to embrace the opportunity to reach today's youth. "We have a huge challenge in front of us," he says. "You as educators have a particularly important job in communicating the basics of climate and the context around it."

Afterwards, one teacher brings up the fact that the film uses the word 'truth' in its title. But isn't science only supposed to deal with theories, he asks, and how does one explain this to students? Another teacher points out that his students didn't understand the meaning of peer-reviewed science as discussed in the film. Isn't it likely that this point is missed by most of the American public, too, he asks. McCaffrey suggests that a role-playing activity on peer review might help drive the point home.

After all, not all teenagers are as attentive to the science of climate change as 14-year-old Alex Budd, of Boulder, who believes that doing something about climate change is a moral imperative. Alex heard about the programme from an aunt who lives in Tennessee, and was the youngest volunteer trained under Gore's tutelage. He and McCaffrey, along with another Colorado-based volunteer, Steve Wilton, have given the presentation together in the Boulder area. "We're destroying our planet," Alex says matter-of-factly. "That's not an issue of politics or economics. That's just morally wrong." He brings up what he learned from the Gore training over lunch at school "or anywhere that I can just say a word or two". The biggest difference can be made by the small things people can do, he says, rattling off a list that includes compact fluorescent light bulbs and improving home insulation.

"I really wanted to make sure people at least know what's happening," Alex says. "It's not going to be easy. That's why they call it an inconvenient truth. It's true but it's not something that fits right into your schedule." ■

Amanda Haag is a science writer in Colorado.

Save your notes, drafts and printouts: today's work is tomorrow's history

SIR — Science is one of the greatest cultural achievements of humankind. And yet — although we assiduously preserve the preparatory sketches of artists, the drafts of novelists and the manuscript scores of composers — there is little systematic preservation of the workings of scientists. This is certainly regrettable for historical studies of modern experimental biology. Since the discovery of the double helix in 1953, biological research has flourished at an ever-increasing pace and many basic insights continue to emerge. Our knowledge of the workings of organisms from all branches of life is increasing at an unprecedented rate, making it imperative that we document the history of these discoveries.

Most recently, the computational analysis of the completely sequenced genomes of many organisms are driving research and guiding experiments. A new generation of tools such as microarrays, advanced imaging systems and single-molecule techniques are fundamentally changing experimental protocols. Where are the original notes, and the patent and manuscript drafts that accompanied these stupendous advances? Nowadays, these are recorded in ephemeral electronic media that are far too easily lost with the push of a button or the failure of a hard drive. Yet historians need all forms of data about the workings of scientists so that they can document the development of today's innovations and inspire future generations to pursue similar lofty goals in science.

Along with these advances in academic science, the new industry of biotechnology came into being. Many of the scientists who led advances in the laboratory were instrumental in establishing biotechnology as a central discipline. Entrepreneurs and venture capitalists also played an important role, recognizing how research in academia could be applied for the benefit of society. Their records, too, will throw an important light on scientific history.

Fortunately, there is increasing interest among historians of science and institutional archives in preserving this history. Top-notch institutions across the United States are establishing archival collections related to the history of molecular biology and chemistry. Taking a lead in this endeavour is Cold Spring Harbor Laboratory, which has recently expanded its library and archives by establishing the Genentech Center for the History of Molecular Biology and Biotechnology (see http://library.cshl.edu/GCHMBB/index.html). This is funded through the generosity of the pioneer biotechnology company, Genentech. Several important collections, including Jim Watson's and S. B.'s personal papers, are already in the archives or pledged for the future.

We encourage all who have played a part in the developments of molecular biology and biotechnology over the past 50 years, and who are continuing this remarkable journey into the future, to preserve their papers and donate them to institutions that are committed to making them freely accessible to scholars. Let's not wait until memories have faded and papers been discarded at the end of a career before deciding to save our heritage. Future historians of science and social science should not have to look back and wonder how it was possible that we discarded the records of our lives in science.

Sydney Brenner*, Richard J. Roberts†
*Salk Institute for Biological Studies, PO Box 85800, San Diego, California 92186, USA
†New England Biolabs, 240 County Road, Ipswich, Massachusetts 01938, USA

Reliance on bibliometric databases can let you down

SIR — Publications not indexed in listings such as the ISI Web of Science are, these days, considered of questionable merit. In more collegial times, research performance not adequately represented by application of such standardized metrics could be evaluated fairly — for example, with allowances for lack of coverage of some disciplines, for citation behaviour in different disciplines, and for the existence of prestigious alternative forums. Your News Feature "The counting house" (*Nature* **415,** 726–729; 2002) drew attention to some problems with bibliometric databases and their uses, and many of the 64 citations of this News Feature listed since then in the Web of Science provide further analyses of problems.

ISI has recently delisted a number of publications from the Web of Science without informing the affected publishers or editors, or publishing a full list of the excisions. The motivation seems to have been to focus the Web of Science on journals and to move conference proceedings to another, little-known product, ISI Proceedings — notwithstanding the fact that many journals have special issues containing conference proceedings.

Proceedings of the Combustion Institute, an important archive in the multidisciplinary field of combustion dating back to 1928, is one of the affected publications. Because its peer-reviewed papers are presented at the biennial International Symposium on Combustion, they will no longer be listed in the Web of Science. According to ISI, the decision to exclude this publication "was not based on an evaluation of its importance to the community of scholars it serves".

This experience adds a new dimension to problems with excessive reliance on citation analyses. The Web of Science database itself is subject to unaccountable adjustments without scientific justification or regard to scientific importance.

Brian Haynes
The Combustion Institute, 5001 Baum Boulevard, Suite 635, Pittsburgh, Pennsylvania 15213-1851, USA

Increasing prose quality by decreasing word repetition

SIR — 'Increase' and 'decrease' are serviceable English words, so why is it my mission to winnow them from the prose that I edit daily? As a technical editor in a university department, I do not demand poetry from my writers; scientific accuracy and logical flow are paramount. Nevertheless, I long for an occasional fresh alternative to 'increasing' and 'decreasing' quantities, measurements and all manner of other too-familiar turns of phrase.

Must mice always have 'a decreased tail length'? I admire the professionalism that refrains from a description of 'adorable, stumpy little mouse tails', but what is wrong with 'shorter tails'? It saves two words for writers tearing their hair out over journals' word counts, and is no less precise. 'Fluoresce' is a lovely word, so why ruin its inherent lyricism with a dull 'increase'? Try 'brighter' fluorescence occasionally, or even 'more intense'.

I challenge all scientific authors: search your documents and count how often you use these two simple words, not forgetting permutations such as 'increasing' and 'increased'. You may be surprised at how frequently they rear their heads.

If so, I urge you to seek a remedy. There are times when only an increase or a decrease will do. Make those times count, and use the full expanse of the English language to broaden your prose elsewhere. Sheer repetition is anaesthetizing, and the aim (one hopes) is to keep the reader awake as well as informed. Strive for accuracy, logic and truth; but in matters of style, simple variety is a welcome spice.

Cheryl Strauss
Department of Human Genetics, Emory University School of Medicine, 301 Whitehead Biomedical Research Building, 615 Michael Street, Atlanta, Georgia 30322, USA

Contributions to Correspondence may be submitted to correspondence@nature.com. They should be no longer than 500 words, and ideally shorter. They should be signed by no more than three authors; preferably by one. Published contributions are edited.

COMMENTARY

LEANDER/ALAMY

Spring-time for sinks

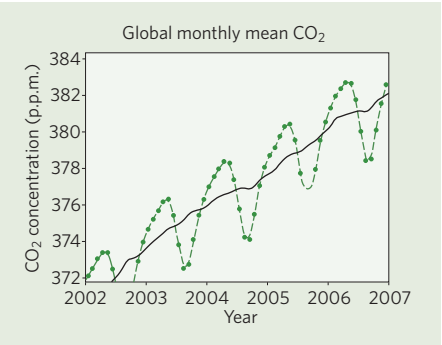
Carbon sinks play a key role in slowing the growth of carbon dioxide levels in the atmosphere. These sinks are at risk as the world warms, but their demise is not inevitable, say **Dave Reay** and his colleagues.

Across the Northern Hemisphere spring is creeping northwards. On the trees the buds are bursting, their leaves unfolding to luxuriate in an atmosphere more enriched in carbon dioxide than at any time in the previous 650,000 years. In the next few months, global CO_2 concentrations will start declining from their annual — and record-breaking — peak[1] (see graph). This happens as the northern biosphere converts more and more atmospheric CO_2 into carbon for storage in terrestrial sinks, such as plants and soils.

With longer growing seasons and earlier spring 'greening', the amount of carbon accumulating in the terrestrial biosphere increased from 0.2 ± 0.7 billion tonnes per annum in the 1980s to 1.4 ± 0.7 billion tonnes in the 1990s. But if the terrestrial carbon sink has played a part in slowing the rate of recent CO_2 increase, the oceanic carbon sink — which mops up twice as much anthropogenic carbon as the land sink — has been even more crucial. Together, these sinks sequester around half of the 6.4 billion tonnes of anthropogenic carbon emitted each year. As such, they are key to determining how rapidly atmospheric CO_2 concentrations, and so global temperatures, will increase during the twenty-first century.

What does the future hold for these sinks? The latest evidence, summarized by the Intergovernmental Panel on Climate Change (IPCC) Working Group II report (WGII) last week, suggests that we cannot rely on the terrestrial and oceanic sinks to go on mopping up our excessive CO_2 emissions indefinitely. As reported by the IPCC, global warming tends to reduce both land and ocean uptake of atmospheric CO_2, so increasing the fraction of anthropogenic emissions that stays in the atmosphere. This is serious because if we take the globally averaged CO_2 concentrations since 1980 and add to these the additional CO_2 that might have accumulated in the atmosphere in the absence of the land and oceanic carbon sinks, we estimate 10% higher CO_2 concentrations in 2006.

Of course, the uncertainties surrounding the impacts of climate change on these carbon sinks are large, but in our view that is no excuse for inaction. By exploring ways in which we can enhance carbon sequestration, we can make the land (and perhaps even the ocean) sinks part of the solution to human-induced climate change, rather than part of the problem.



Sink to source

Given its importance, it is surprising that the impacts of climate change on the huge oceanic carbon sink received scant coverage in the latest WGII report. Uncertainty is doubtless to blame, even though some effects are well understood. As long as CO_2 is increasing in the atmosphere, the ocean will continue to take up carbon. But as CO_2 concentrations rise, oceanic pH will fall, and with the changing chemistry of the ocean, the efficiency with which it absorbs carbon is projected to decrease.

There are other indirect ways that climate change could affect the future oceanic sink. Some changes, such as rising seawater temperatures and increased stratification — whereby less mixing between layers may starve surface waters of nutrients — could reduce CO_2 uptake. Others, such as increased CO_2 uptake in the high-latitude ocean due to less sea-ice cover and increased biological productivity, could increase oceanic uptake. Still, most cli-

mate models predict that a smaller fraction of anthropogenic CO_2 emissions will be absorbed over time as the ocean acidifies[2].

Worse may be in store for the land sink. According to the WGII report, terrestrial ecosystems are highly likely (more than a 90% chance) to become net sources of CO_2 in the latter half of the twenty-first century. This is the strongest statement yet on terrestrial carbon, and has huge implications for global climate change. According to the IPCC's latest predictions, the decline in the oceanic sink and a switch from carbon sink to source on land could increase the global temperature in 2100 by more than 1 °C.

The future of the terrestrial carbon sink may seem bleak, but it doesn't have to be. Yes, the elevated global temperatures and changes in precipitation predicted for 2050 could induce a switch from sink to source as warmer soils release more carbon and forests suffer more drought. In reality, things are more uncertain. Humans have been managing terrestrial ecosystems for our own ends for millennia — from deforestation and agriculture, to increasing fertilizer and pesticide use. Of the 50 billion tonnes of carbon currently locked up in terrestrial biomass and vulnerable to release in the next 20 years, 40 billion tonnes is put at risk not by changes in climate, but by changes in land use. Climate change may be the greatest threat to this huge carbon stock towards the end of the century. In the shorter term, it is chainsaws and ploughs, not drought and extreme temperatures, that we must address.

V. R. CAIVANO/AP

Carbon sinks are at greater immediate risk from land-use changes than from global warming.

Hope for the future

Some may see our transformation of Earth's land surface as part of the problem, that the combined grip of more than 6 billion people is squeezing the terrestrial carbon sinks from every direction. But in this powerful grip may also lie part of the solution.

Deforestation is the single biggest threat to the terrestrial carbon sink. Over the past 200 years it is thought to have been responsible for 30% of the anthropogenic increase in atmospheric CO_2 concentrations[3]. Slowing deforestation rates could therefore help to stabilize carbon emissions from terrestrial ecosystems[4], as could enhanced carbon uptake through afforestation and reforestation.

The conversion of soils from natural to agricultural use has also led to substantial losses in terrestrial carbon. Throughout history, soils are thought to have lost between 40 billion and 90 billion tonnes of carbon globally through cultivation and disturbance[5]. Again, sensitive land-use practice can help to protect the remaining soil carbon stocks, and perhaps reverse recent trends. Through practices such as no-till agriculture, more efficient fertilizer use, and the planting of vegetation with higher carbon returns to the soil, the soil sink can be significantly enhanced. A 2000 IPCC report[4] concluded that through such approaches a cumulative increase in terrestrial carbon of up to 100 billion tonnes is possible by 2050.

"Protecting or enhancing carbon sinks has received inadequate attention."

Despite their huge potential, protecting or enhancing carbon sinks has received inadequate attention from the international community. In 2001, the Kyoto Protocol members agreed that richer nations could use their carbon sinks to meet their Kyoto commitments to reduce emissions. As part of the Protocol's Clean Development Mechanism (CDM), rich nations were also able to count cuts in emissions achieved through afforestation and reforestation projects in developing countries. But sinks could be used to meet only a small part of their total commitment and, of the more than 570 registered CDM projects at the time of writing, only one — the reforestation of degraded land in China's Pearl River Basin — is actually aimed at sink enhancement.

The 2008–12 commitment period for the Kyoto Protocol is coming to an end and the question of what will follow is the subject of much debate. Terrestrial sinks are time limited and reversible, so do not replace the need for emissions reduction. But to make them part of the wider solution they must be better integrated into future international agreements.

A quick fix?

In general, any approach to mitigating emissions or adapting to climate change needs to consider the potential feedbacks on both terrestrial and oceanic carbon sinks. For example, biofuels are growing in popularity as a direct substitute for fossil fuels because of the hope that they can partly offset carbon emissions. A recent estimate suggests that, by 2030, about 750 million tonnes of fossil carbon emissions could be offset by biofuels each year[6]. But some biofuels, such as corn ethanol, have similar net carbon emissions to fossil fuels[7]. It is also likely that any massive increase in biofuel production would be at the expense of existing vegetation and soil carbon sinks.

Can we do anything to enhance the much larger oceanic carbon sink? Efforts to enhance ocean productivity, and therefore carbon uptake, through iron fertilization have had short-term results in areas such as the Southern Ocean. But visually impressive algal blooms may do little to enhance long-term carbon uptake. To make any difference on a global scale you would need to keep fertilizing, indefinitely, a much larger area of the ocean. Attractive as the idea of a 'quick fix' for human-induced climate change may be, such schemes are expensive and their side effects poorly understood. Perhaps climate change will be so rapid, and its impacts so severe, that such geoengineering will one day be required. Today, though, it is through the protection and enhancement of the terrestrial carbon sink that we can better engineer our future climate. ■

Dave Reay is at the School of GeoSciences, University of Edinburgh, West Mains Road, Edinburgh EH9 3JN, UK; Christopher Sabine is at the NOAA Pacific Marine Environmental Laboratory; Pete Smith is at the University of Aberdeen; and Graham Hymus is at Northern Arizona University.
e-mail: David.Reay@ed.ac.uk

1. Tans, P. *Trends in Atmospheric Carbon Dioxide* www.esrl.noaa.gov/gmd/ccgg/trends (NOAA/ESRL, 2007).
2. Watson, A. J. & Orr, J. C. in *Ocean Biogeochemistry: A JGOFS Synthesis* (ed. Fasham, M. J. R.) 123–143 (Springer, Heidelberg, 2003).
3. Raupach, M. R. *et al.* in *The Global Carbon Cycle. Integrating Humans, Climate and the Natural World* (eds Field, C. B. & Raupauch, M. R.) 526 (SCOPE 62, Island Press, 2004).
4. Watson, R. T. *et al.* IPCC special report on Land Use, Land Use Change and Forestry (2000).
5. Houghton, R. A., Hackler, J. L. & Lawrence, K. T. *Science* **285,** 574–578 (1999).
6. Rose, S. *et al.* Energy Modeling Forum Report www.stanford.edu/group/EMF/projects/group21/EMF21sinkspagenew.htm (Stanford Univ., 2007).
7. Farrell, A. E. *et al. Science* **311,** 506–508 (2006).

BOOKS & ARTS

Industrial relations

What lessons can be learned from a landmark deal between a company and a university department?

Universities in the Age of Corporate Science: The UC Berkeley–Novartis Controversy
by Alan P. Rudy, Dawn Coppin, Jason Konefal, Bradley T. Shaw, Toby Ten Eyck, Craig Harris & Lawrence Busch
Temple University Press: 2007. 256 pp. $54.50

Scott Wallsten

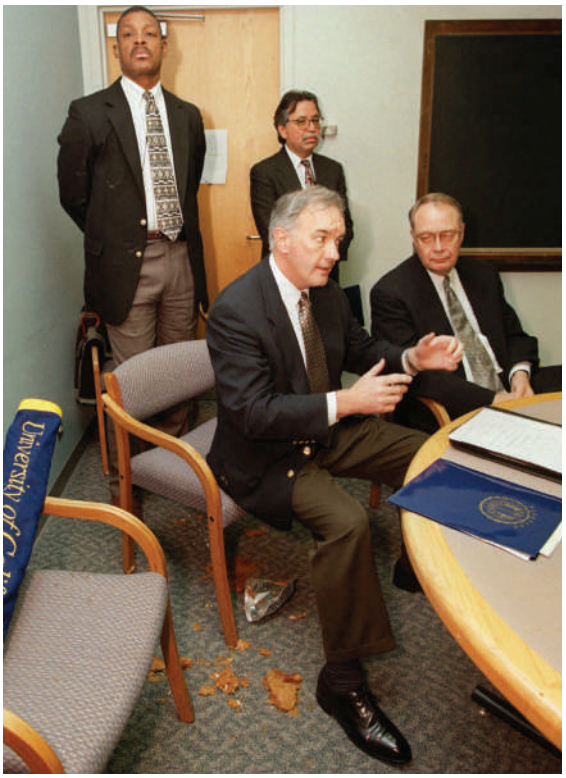
Universities can be ambivalent about their relationships with industry. Ideally, researchers can carry out fundamental scientific research without having to worry about commercial applications. Industry, on the other hand, is less likely to undertake research into basic science because of the limited return, and so is more likely to focus on research that yields commercial products. In principle, then, university and industry research are complementary.

But the reality is not always this simple. Industry scientists sometimes pursue fundamental science, and university scientists sometimes work on projects that will potentially yield commercial payoffs. Moreover, there is not always a clear line between applied and fundamental research. Industry and university research can overlap substantially, with collaborations producing benefits to society as well as supporting financially constrained academics.

The challenge is how to foster relationships between industry and academic scientists in ways that enable university researchers to keep their academic independence while allowing companies to generate returns for their shareholders. Balancing these objectives is not easy.

Industrial support of university research is nothing new, and in recent years both governments and universities have tried to find ways of bringing university research to the market. Nevertheless, the agreement between the University of California, Berkeley, and the pharmaceutical giant Novartis (the UCB–N agreement) in 1998 was unusual in both its size and its scope, and was highly controversial (see *Nature* **399,** 5; 1999). As a way of resolving the controversy, the university agreed to commission a study of the UCB–N agreement and its impact. *Universities in the Age of Corporate Science* is the outcome of this study.

The book provides fascinating details of the deal, the players and the controversy, and does an admirable job of empirically and qualitatively measuring the effects of the agreement on scientific research. It succeeds in its aim of analysing the UCB–N deal despite being written in a way likely to appeal more to sociologists than to people interested specifically in the issue, especially in the early part of the book — the reader is not properly introduced to the UCB–N agreement until the fourth chapter.



G. NIKITIN/AP

Crushed underfoot? Many researchers were unhappy at a deal between the University of California, Berkeley, and Novartis.

Alan Rudy and co-authors attempt to put a theoretical structure on the analysis — a laudable intention as testable theories help to frame analyses and clarify the precise questions being asked. In this case, though, theory does little to structure the analysis. The reader's reward for ploughing on is an in-depth discussion of the UCB–N deal and the firestorm it ignited.

The authors explain the many factors that came together to intensify the controversy. First, the deal was unique in that the company made the agreement with an entire university department, rather than with just a single researcher. As a result, and unlike the usual grant funding system, the agreement subjected an entire population of researchers to its stipulations. It also seems to have exacerbated some resentment between departments over resources. Second, many perceived that the deal was not sufficiently transparent and that it was made with scant input from groups that could have been affected, such as graduate students and postdocs. Third, Novartis was given unprecedented access to the department's research, including the right of first refusal to licensed discoveries — regardless of whether or not they were funded by Novartis.

Ultimately, though, none of the opponents' fears about industry affecting university research were realized. The authors conclude: "it appears that UCB–N resulted in modest benefits and very little harm" to the department.

The authors labour to draw general lessons from the UCB-N experience, although it is not obvious that there are many to be drawn. Generalizing lessons from a case study is often problematic, particularly if there are factors specific to the case. For example, as the authors note, Novartis was undergoing radical corporate restructuring throughout much of the agreement, so the extent of its involvement may have been affected, even though its funding commitments were met.

Case studies are typically chosen precisely because there is something unusual about them. The furore surrounding the UCB–N deal also makes it hard to generalize. Potential abuse, for example, may have been kept in check while people were keeping a close watch on the collaboration, and interactions between scientists from both sides may have been affected for fear of fanning the flames. In other words, both the benefits and the costs of the agreement may have been muted by factors unique to the UCB–N agreement.

Complaints would not necessarily have been resolved more satisfactorily if the university had handled them differently. Graduate students and postdocs felt left out of the process, but what should their level of involvement have been? The interests of both groups are short term compared with those of the faculty. The outcome may have been adversely affected if

too many parties were involved, as negotiations could have become unwieldy and agreement less likely.

Collaborations between university and industry can bring large benefits: more resources to university faculty and students, faster dissemination of research results, new products on the market, and fresh insight from interactions that might not otherwise have happened. But such collaborations can be difficult to manage because universities and industries have different objectives.

In short, *Universities in the Age of Corporate Science* is a compelling and detailed description of the events surrounding the UCB–N deal. It should be enjoyed by all those who follow the evolution of university–industry relations, offering as it does a unique look at how the collaboration was made. ■

Scott Wallsten is a senior fellow at the Progress & Freedom Foundation, 1444 Eye Street, NW, Washington DC 20005, USA.

Builders with little brains

Animal Architects: Building and the Evolution of Intelligence
by James L. Gould & Carol Grant Gould
Basic Books: 2007. 316 pp. $26.95

Tore Slagsvold

Running, climbing, diving and surviving in extreme environments are just some of the physical skills in which animals excel compared with humans. But we are confident that our bigger brains make us better at tasks demanding intelligence. Take building, for example — we can build cars, houses and castles. But to what extent is intelligence actually needed to be a good architect and builder, and are humans really unique in this respect?

These questions are raised by James Gould, professor of ecology and evolutionary biology at Princeton University, and science writer Carol Gould in their book *Animal Architects*. They point out that tiny termites can build a tower 6 metres high. Taking relative body length into account, this would be equivalent to about a height of about 4 kilometres for a human, making the Eiffel Tower and the Empire State Building seem very small indeed. Even in absolute size, coral reefs — the largest structures built by animals and the only signs of life on our planet that are visible from space — are beyond the scope of human creations.

In this book, the authors show how spiders build webs with silk to catch their prey; how a silkworm weaves its cocoon around itself; how honeybees use wax to build precise hexagonal combs for their young and for honey; how paper wasps masticate materials from twigs with water to make damp cellulose for their combs; and how insects get air-conditioning for their homes. Birds are also sophisticated builders of nests. But building skills are generally less developed in mammals, as the safety of the womb renders additional prenatal protection redundant — although there are some notable exceptions, such as the beaver.

Does building call for great cognitive ability? Not necessarily, say the authors, who use knowledge from ethology about sign stimuli, motor programmes and motivation to explain how impressive constructions can be built from many small steps. This explanation seems plausible — after all, I can cook a complex meal by following a detailed recipe, and I may even be able to build a small cabin.

The question of intelligence is integrated into the whole of the book, which is a notable achievement. The authors even offer a provocative analysis of one of our own skills that we consider to be very advanced — language learning — and show that it is based on a simple chain of built-in recognition systems for sign stimuli and on innate motor programmes. Such innate behaviours do not preclude the existence of higher cognitive processes, however: indeed, these allow us to perform largely on 'autopilot' while focusing on whatever cognitively challenging task arises.

We also believe we have a unique aesthetic sense, yet some animals, such as bower birds, build small huts decorated with colourful objects. The bower's only purpose is to help attract a mate. Bower birds are considered to be intelligent, suggesting that recursive cycles of selection for a single set of cognitive building abilities and aesthetic refinements are part of the same sort of positive-feedback loop that may have led to the evolution of the human mind. Well, parrots and crows are smart too, but most parrots do not build nests, and crows make only simple nests from sticks (albeit with a lining). The evolution of bigger brains may have more to do with sociality and the necessary communication skills. Humans have no impressive evolutionary past as builders and originally lived in simple caves. But we used tools for hunting and fighting, and the authors may be correct in saying that such skills contributed to brain evolution.

The illustrations in this highly readable book are in black and white, which does not do full justice to the animal architects' remarkable achievements. Each of the ten chapters has a separate list of literature for further reading, but as references are not included in the text it is hard to figure out what is new and which contributions are the authors' own.

Animal constructions are fascinating, and the authors provide some useful insights into them. They show how the creation of complex constructions depends on evolutionary history and the investment of time and energy. Although the builders may not be particularly intelligent, their buildings serve their purpose well and confer fitness benefits on the architects. Could it be that our own building activity is driven not only by the need to shelter from the storm, but also by the desire for power and mate attraction? ■

Tore Slagsvold is in the Department of Biology, University of Oslo, Blindern, N-0316 Oslo, Norway.

M. & P. FOGDEN/FLPA

Attractive design: male bower birds build nests adorned with colourful items in a bid to lure a mate.

The birth of science

The Emergence of a Scientific Culture: Science and the Shaping of Modernity 1210–1685
by Stephen Gaukroger
Oxford University Press: 2006. 576 pp.
£35, $65

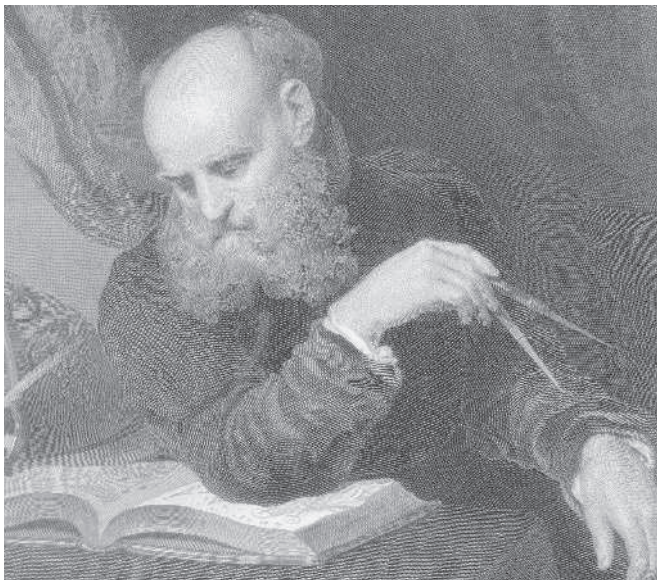
Peter Dear

Why do we look to science to guide our lives in the modern world? That most of us do so is manifest from an examination of the cultural and political history of the past century. This point is made very early on by Stephen Gaukroger in his new book *The Emergence of a Scientific Culture*, when he quotes Jawaharlal Nehru. But what does it mean, and how, historically, did it come to be the case? Not surprisingly, Gaukroger's attempts to resolve such a grand issue are, literally, weighty, as this volume constitutes the first of a projected five. It is not an easy question to answer.

By starting the story in thirteenth-century Latin Europe, Gaukroger presents a world in which theology, not 'natural philosophy', was regarded as the 'queen of the sciences'. He then traces how this gradually ceased to be the case, and natural philosophy, albeit of a new kind, displaced theology as the touchstone of cognitive propriety. By the end of the seventeenth century, many people were arguing that the standards and procedures of natural philosophy were appropriate models for all kinds of cognitive enquiry, including those involving theology and religion.

Gaukroger's big claim is that, although technical scientific endeavours and achievements have cropped up in various world cultures at different times, it was only in the European culture that formed in the seventeenth century that these advances became a continuous tradition. What made Europe different is that a new form of philosophical enquiry into nature took the lead in determining what were still the culturally dominant questions in Europe, namely those concerning religion. Far from separating itself from religion, European science at the time became the principal tool for underpinning it: understanding nature was the path towards knowledge of God. And having attained this status by the 1680s, argues Gaukroger, science hasn't looked back since.

There are many complexities in the book's overall narrative. One is the relationship between modern science, which as a coherent, institutionalized endeavour is usually seen by historians (including Gaukroger) as having emerged in the nineteenth century, and the earlier field of natural philosophy. Natural philosophy since the thirteenth century had been centred on the works of Aristotle, and was defined as a contemplative intellectual pursuit aimed at providing causal accounts of natural phenomena. It explicitly eschewed practical ends, seeking only cognitive understanding.



MARY EVANS PICTURES

Galileo was central to the rise of a mathematical 'natural philosophy'.

The great changes of the seventeenth century that Gaukroger details involved a rejection of many features of aristotelian natural philosophy. In its place appeared a kind of natural philosophy that incorporated practical mathematical sciences (such as astronomy, mechanics and optics) and forms of 'mechanism' that offered explanations of macroscale phenomena in terms of the underlying structure of matter (atomistic or corpuscularian).

Figures such as Francis Bacon in the early seventeenth century further reconstructed the mission of natural philosophy by rejecting its aristotelian restriction to contemplative knowledge. They argued instead, in explicitly religious terms, that its proper role was precisely to produce control over nature for the practical benefit of mankind. Indeed, Gaukroger here makes a powerful case that this development formed part of a larger reconstruction of natural philosophy that related it more closely to moral philosophy. Other ancient Greek philosophers, he argues, had understood their task as one of determining how best to live in the world, with ideas about nature being subservient to that end.

In rejecting Aristotle, seventeenth-century natural philosophers revitalized that old ideal, and in doing so adopted a new persona, placing a premium on objectivity and trustworthiness that outshone dogmatic claims to knowledge of truth. The experimental philosophy of Robert Boyle and Isaac Newton, for example, set itself up as a model of disciplined enquiry to be emulated by the rest of society, rather as experimental philosophy owed much to legal practices. Whether the natural philosophers of the seventeenth century, with their great variety of commitments and approaches, should therefore be seen as scientists *avant la lettre* is a question that Gaukroger in effect defers.

Despite its sweeping chronological scope, Gaukroger's book focuses largely on prominent figures of the seventeenth century, including Johannes Kepler, Bacon, Galileo and René Descartes. Gaukroger provides an insightful analysis of their natural-philosophical ideas and arguments, paying special attention to the epistemological conceptions that each represented and developed.

The book's title uses the term 'scientific culture', but its content also reminds us of its author's accomplishments as a historian of philosophy. Why the changes that Gaukroger describes should have occurred in seventeenth-century Europe remains largely unanswered, but this book contributes to the formulation and solution of that possibly intractable question, and promises even more in future volumes. ■

Peter Dear is in the Department of History, Cornell University, Ithaca, New York 14853, USA.

SCIENCE BOOKS AWARD

Judging for the Royal Society Prizes for Science Books is well under way. The longlist of books for the general prize includes a wider mix of popular science books than in previous years, with **Giant Leaps** by journalist John Perry of UK newspaper *The Sun* and Jack Challoner vying with Nobel laureate Eric Kandel's memoir **In Search of Memory**. The mysteries of the mind are also explored by Cordelia Fine in **A Mind of its Own** and Daniel Gilbert in **Stumbling on Happiness**. Adam Wishart's **One in Three** is a personal exploration of the biology and treatment of cancer. Current issues are represented by Robert Henson's **The Rough Guide to Climate Change** and TV tie-ins by **The Science of Doctor Who** by Paul Parsons. Established writers Matt Ridley (who wrote **Francis Crick**), Paul Davies (**The Goldilocks Enigma**) and Chris Stringer (**Homo Britannicus**) appear alongside newcomer Henry Nicholls (**Lonesome George**). Rock guitarist Brian May joined experienced astronomers Patrick Moore and Chris Lintott to write **Bang!**

This list will be whittled down to a shortlist of six, and the winners of both the general and the junior prizes will be announced on 15 May 2007.

www.royalsoc.ac.uk/sciencebooks

Capturing human behaviour

Understanding the dynamics of infectious-disease transmission demands a holistic approach, yet today's models largely ignore how epidemics change individual behaviour.

Neil Ferguson

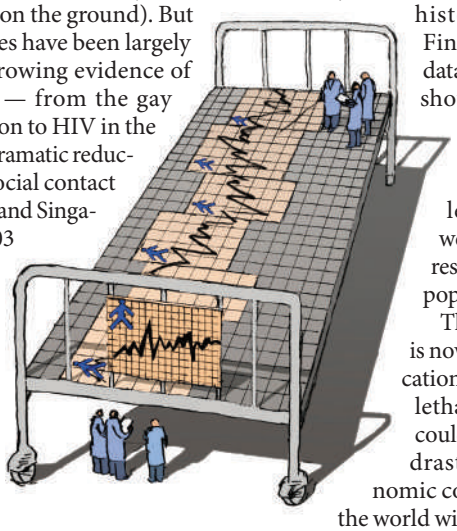
We live in an ever more connected, mobile and interdependent world, where small perturbations can have unpredictable and sometimes far-reaching effects. The paradox is that we increasingly demand predictability. From climate to car design, we expect the future to be anticipated, risks assessed and solutions to be rational. We have to be 'on top' of everything — including threats from infectious diseases.

In response to this trend, policy-makers increasingly turn to epidemic models as a tool in tackling potentially catastrophic outbreaks, from Britain's foot-and-mouth epidemic of 2001 and SARS in 2003 to the pandemic threat now posed by the H5N1 strain of avian flu. Better data, significantly boosted computer power and theoretical advances — particularly from the social sciences — have endowed models with a new realism. Yet fundamental limitations remain in how well they capture a key social parameter: human behaviour.

The subtlety in epidemic modelling lies in how the processes of human contact that underlie transmission and the biology of the host–pathogen interaction are represented. It is in this area that social science has had the greatest impact. Past models represented societies as 'compartments' of identical individuals all mixing randomly; the new paradigm is social networks characterized by either casual or intimate contact — the former being more relevant for respiratory diseases, the latter for sexually transmitted infections.

This is an improvement, but a limited one. The problem lies in the highly simplified way the description of disease and transmission is layered onto the social network, which gives a picture that is both biologically and socially flawed. Most glaringly, the effects of behavioural responses to epidemics are given short shrift. The social networks in most models are usually static, allowing scant interaction between the epidemic and individual or group behaviour, bar sick people staying at home — and not all models include that. For mild infections this level of sophistication may be justified, as we rarely fundamentally change our behaviour because of a cold; but increasingly, the evidence shows that faced with lethal or novel pathogens, people will change their behaviour to try to reduce their risk.

Organized measures such as quarantine, contact tracing or closure of public places are one way that behaviour can change, but people may also spontaneously modify their behaviour to reduce perceived risk. Public-health measures are often examined prospectively in models (although many retrospective studies of outbreaks only crudely capture the complexity of controls implemented on the ground). But individual responses have been largely ignored, despite growing evidence of their importance — from the gay community's reaction to HIV in the early 1980s, to the dramatic reduction in travel and social contact seen in Hong Kong and Singapore during the 2003 SARS epidemic. Even absenteeism resulting from illness or caring for sick dependants can significantly affect close-contact networks, by removing people from workplaces.



J. KAPUSTA/IMAGES.COM

By modifying the contact network, behavioural changes during an epidemic can give dynamics very different from the kind predicted by simple models. Most basic models assume that all parameters are static, but in fact people's responses often shift as an epidemic progresses. Individuals are most likely to change their contact patterns when mortality or the perception of risk is high, and resume normal life as the perceived risk declines. A case in point is the recent resurgence of risky behaviours in some gay communities following the widespread availability of combination antiviral therapy for HIV; another, the ongoing studies of how public-health measures and individual risk-reduction behaviour shaped the very different epidemic patterns seen in various US cities during the Spanish flu pandemic of 1918–19.

The challenge for mathematical modellers is that data are scarce, and often qualitative when they do emerge. An example from the UK foot-and-mouth disease epidemic is anecdotal evidence of more people failing to comply with biosecurity restrictions around farms once the epidemic was in decline, which may have contributed to that outbreak's long tail.

Bridging this data gap, and developing succinct yet realistic descriptions of epidemics' impact on social-network dynamics, will prove key to making models of lethal epidemics more accurate. Three avenues of research need to be pursued. The first is controlled epidemiological intervention studies to determine how modifying contact networks affects disease transmission. Another will be the integrated collation and analysis of epidemiological and, preferably, quantitative social data from historical epidemics. Finally, protocols and data-collection systems should be designed to track the number of people infected per day in a future lethal outbreak as well as the behavioural response of the affected population.

The time for this work is now: global communications mean that a novel lethal disease outbreak could trigger potentially drastic social and economic consequences across the world within days.

From a public-health perspective, the goal is improving our ability to predict and control epidemics — but that may first require new sociological models that are both predictive and quantitative. So the interdisciplinary approach remains vital, this time at the interface of epidemiology, sociology and the history of medicine.

Beyond public health, what is there in this enterprise to motivate sociologists, anthropologists or historians? Understanding behavioural responses to lethal infectious diseases may help epidemic modelling now, but over time it could reshape our understanding of the interaction between disease and society as one of coevolution. Historically, certain diseases have exploited social upheavals such as urbanization, and behavioural responses to disease risk — such as the hand-use bias in Indian dining habits — have become part of the cultural fabric. But we have barely begun to unravel the larger question of how disease has shaped behavioural norms and, through those, society as a whole. The answers may be surprising. ■

Neil Ferguson is at the MRC Centre for Outbreak Analysis and Modelling, St Mary's Campus, Imperial College London, London W2 1PG, UK.

For other essays in this series, see http://nature.com/nature/focus/arts/connections/index.html

NEWS & VIEWS

CANCER

Division of labour

Gerhard Christofori

Some genes are involved in the development of a new tumour; others specifically promote the dissemination of its cancerous cells to other organs. A set of four genes seems to be required for both processes.

Most deaths from cancer are due to metastasis, or the spread of cancer cells. The discovery of gene sets, or 'signatures', whose expression is associated with general and organ-specific metastasis has prompted the first screening programmes in the clinic[1,2]. Together with the identification of a growing number of genes involved in several steps of the metastatic process[3], the search for molecular targets that are amenable to therapy is a major focus of cancer research. On page 765 of this issue, Gupta *et al.*[4] report that four of the lung metastasis signature genes mediate a "multifunctional vascular remodelling programme".

Researchers from the same laboratory had previously generated variants of the breast cancer cell line MDA-MB-231 that exhibit organ-specific metastasis to either bones or the lungs[5–7]. They identified bone- and lung-specific signatures of metastasis through analysis of gene-expression profiles, and confirmed the role of some of these genes in organ-specific metastasis. Gupta *et al.*[4] have now carried out experiments involving genetic and pharmacological manipulation of the function of these genes in tumour cells. They report that four of the genes promote not only the growth of primary tumours but also the entry of cancer cells into the vasculature (intravasation), their colonization of the lung, permeation through blood vessels (extravasation) and metastatic outgrowth.

The products of these four genes — epiregulin, two metalloproteinases and cyclooxygenase-2 (COX-2) — are well known to cancer researchers. Epiregulin is a ligand for the epidermal growth factor receptor (EGFR) and is essential for the growth, survival and progression of several types of cancer. The matrix metalloproteinases (MMP-1 and MMP-2) participate in the formation of new blood vessels to supply tumours (angiogenesis), as well as in tumour-cell migration and invasion. The COX-2 enzyme mediates wound healing and inflammatory responses, and has previously been implicated in promoting metastasis[8].

Gupta and colleagues[4] found that when any of these genes is inactivated by genetic manipulation in breast cancer cells that are metastasizing to the lung, there is only a moderate inhibition of primary-tumour growth and lung metastasis. But when combinations of these genes are inactivated, additive or synergistic effects are apparent, with an almost complete abrogation of both primary-tumour growth and lung metastasis when all four genes are inactive. Although proliferation is unaffected in tumour cells in which all four genes are inactivated, apoptosis — programmed cell death — is increased. Moreover, the blood vessels of such tumours have smaller lumens and are less permeable, factors that are probably responsible for the increased apoptosis and the failure of the tumour cells to enter the bloodstream and extravasate in the lung. Indeed, tumour cells with all four genes inactivated are trapped within capillaries of the lung, whereas their normal counterparts nested outside the vessels in the parenchymal tissue of the lung to grow as a 'micro-metastasis'. The authors[4] confirmed the inability of genetically manipulated tumour cells to permeate the endothelial barrier — the cell layer covering the luminal side of blood vessels — in a migration assay.

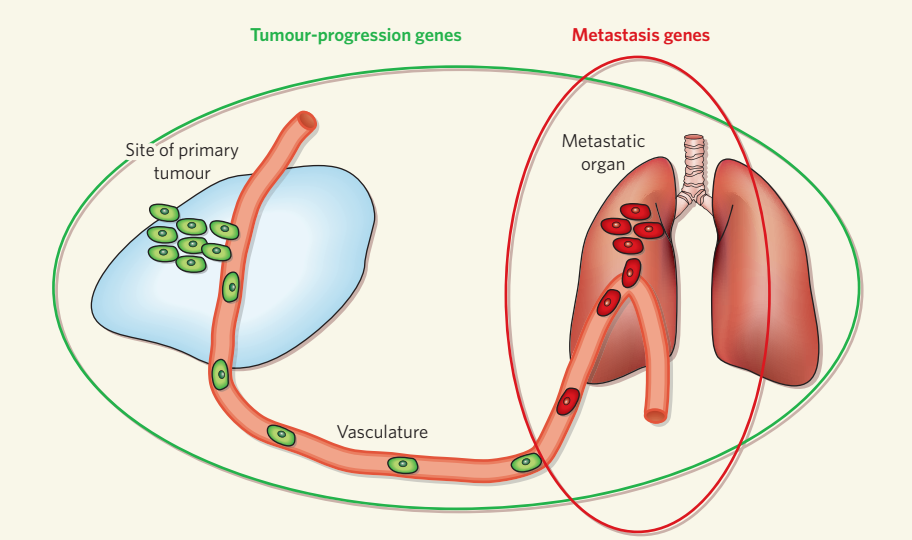


Figure 1 | Roles of different genes in cancer. Gupta *et al.*[4] identify genes that are required for the growth of primary tumours, for intravasation and extravasation of tumour cells to the specific metastatic target organ (as exemplified here by the lung) and for metastatic outgrowth. These genes are required for the growth of the primary tumour and the many stages of metastasis, and can be classified as 'tumour-progression genes'. By contrast, 'metastasis genes' contribute exclusively to metastatic outgrowth.

These data[4] indicate that epiregulin, COX-2 and MMP-1/2 act synergistically to remodel the vasculature in primary tumours by forming a tortuous and highly permeable network of blood vessels. Such a network is often observed in aggressive tumours. These proteins seem to have a crucial role in both the intravasation of tumour cells and the breaching of the endothelial barrier in the metastatic organ to allow their extravasation. By contrast, combinatorial ablation of signature genes that exclusively affect lung metastasis, such as *IL13Ra2*, *SPARC* and *VCAM1* (ref. 6), did not affect primary-tumour growth[4]. Thus, there seems to be a division of labour between different gene categories — metastasis genes are exclusively involved in metastasis, and not in primary-tumour outgrowth, whereas tumour-progression genes function in both (Fig. 1).

Gupta *et al.*[4] also used pharmacological agents to specifically inhibit the products of the four genes. Cetixumab, a neutralizing antibody

against EGFR, and celecoxib, a specific inhibitor of COX-2, are in clinical use, and GM6001 — a broadband MMP inhibitor — has been tested in preclinical trials. Combinatorial treatment of mice transplanted with the lung-metastatic variant of the MDA-MB-231 cell line recapitulated the results of the genetic ablation studies. The authors observed efficient repression of primary-tumour growth, as well as tumour-cell intravasation, extravasation and lung metastasis. This treatment also inhibited lung metastasis in mice intravenously injected with primary breast cancer cells and their derivatives, the latter of which express high levels of epiregulin, COX-2 and MMP-1. Cessation of treatment resulted in the release of tumour cells trapped within the lung capillaries into the lung parenchymal tissue.

These studies demonstrate the functional dissection of the metastatic process and identify specific genes involved in its many stages. The genetic and pharmacological manipulation generated the same outcome, indicating that pharmacological treatment affects the activities of the four metastasis-progression genes only in tumour cells and not — as has been assumed — in cells of the tumour microenvironment, such as COX-2 in endothelial cells or MMP-1/2 in tumour-infiltrating inflammatory cells.

The findings of Gupta *et al.* also have bearings on a question that has intrigued cancer researchers for many years. Are the events that modify genes — either directly (genetic) or indirectly (epigenetic) — and that are required for metastasis, already predetermined in primary-tumour cells or do certain cancer cells gain metastatic ability in a 'micro-evolutionary' selection process[9]? The authors argue that some genetic and epigenetic events affect genes that are already present in the primary tumour and are required for metastatic outgrowth (tumour-progression genes), whereas other such events affect genes that are required only for the metastatic process (metastasis genes; Fig. 1). How and whether these genes are co-selected by function remain unclear.

Further studies are now warranted into the molecular mechanisms that underlie the remodelling of tumour vasculature and its consequences. Which other cell types and factors are involved in the remodelling process? How is organ specificity determined? What are the chemoattractants and adhesion molecules that target the cancer cells to the lung? It has been reported that bone-marrow-derived progenitor cells are recruited to the lung to set up a metastatic 'niche' for disseminating cancer cells[10]. In addition to their role in vascular remodelling, are these cells involved in the organ-specific metastasis observed here? Is the combined activity of epiregulin, COX-2 and MMP-1/2 in the extravasation process specific to lung metastasis, or does it also occur in other target organs? Answers to these questions will also add long-sought molecular details to Paget's 'seed and soil' hypothesis[11], which states that disseminating tumour cells (seed) need to find the appropriate microenvironment in distant organs (soil) for metastatic outgrowth.

Other questions concern the implications of Gupta and colleagues' study[4] for cancer therapy. Are the four genes investigated here critical players in all subtypes of breast cancer, and possibly in other types of cancer as well? Can the combinatorial treatment, which has proved so successful in the preclinical setting reported by these authors, be further developed for clinical application?

Finally, on a technical note, the work of Gupta *et al.* marks a technological milestone. The ability to generate stable cell lines that carry four genes with suppressed expression at the same time, and the simultaneous forced expression of three of these four genes to test for specificity of gene ablation, would be an impressive experimental feat. When used more widely, it will be instrumental in elucidating the combined function of genes during cancer development. ■

Gerhard Christofori is at the Institute of Biochemistry and Genetics, Department of Clinical–Biological Sciences (DKBW), University of Basel, Center for Biomedicine, Mattenstrasse 28, CH-4058 Basel, Switzerland.
e-mail: gerhard.christofori@unibas.ch

1. Buckhaults, P. *Curr. Opin. Oncol.* **18,** 57–61 (2006).
2. Segal, E., Friedman, N., Kaminski, N., Regev, A. & Koller, D. *Nature Genet.* **37,** S38–S45 (2005).
3. Christofori, G. *Nature* **441,** 444–450 (2006).
4. Gupta, G. P. *et al. Nature* **446,** 765–770 (2007).
5. Kang, Y. *et al. Cancer Cell* **3,** 537–549 (2003).
6. Minn, A. J. *et al. Nature* **436,** 518–524 (2005).
7. Minn, A. J. *et al. J. Clin. Invest.* **115,** 44–55 (2005).
8. Roh, J. L. *et al. Cancer Res.* **64,** 3230–3235 (2004).
9. Bernards, R. & Weinberg, R. A. *Nature* **418,** 823 (2002).
10. Kaplan, R. N. *et al. Nature* **438,** 820–827 (2005).
11. Paget, S. *Cancer Metastas. Rev.* **8,** 98–101 (1989).

SOLID-STATE CHEMISTRY

Crystal tennis rackets

J. Michael McBride

The idea of bendy crystals, especially ones that move rapidly and reversibly in response to light, seems strange. Such materials have now been prepared — but how do they change shape so dramatically without cracking?

Most chemical reactions are accompanied by a change of molecular shape. Impressively, evolution has harnessed this effect to generate mechanical power for biological 'devices', for example, in muscles, where the bending of myosin upon release of adenosine diphosphate is used to drive one fibre past another[1]. In stark contrast, science has produced very few artificial analogues. In most cases, chemical transformations contribute to man-made mechanical devices only at the crudest level, by supplying a source of heat, or electrical energy, or indiscriminant swelling. Current work on nanodevices is changing this situation at the molecular level[2], but problems in linking such devices together to achieve larger-scale motion have inhibited progress. On page 778 of this issue, Irie and colleagues[3] describe a breakthrough in this area — a light-induced chemical transformation that bends crystals without shattering them.

So how does one go about exploiting molecular shape-changing to generate movement at the bulk level? For practical applications, the chemical transformation should be reversible — both chemically and mechanically — and the interconverting molecular states should generally be stable. Numerous reversible, light-induced reactions might fit the bill, possibly providing the basis for actuator devices that convert light into motion. For example, molecules can be constructed that have rigid double bonds at their centres. But when these compounds are irradiated with ultraviolet or visible light, rotation about the double bond becomes possible, and the molecules can change shape. Similarly, certain chains of six carbon atoms can form a ring upon irradiation, so as to flatten a previously bulky molecule. This process can sometimes be reversed by changing the wavelength of the irradiating light, and it is this sort of system that is used by Irie and colleagues[3].

But for practical purposes, there is another requirement: the molecules concerned must be ordered and coupled so that millions or billions of them cooperate to achieve large-scale mechanical motion. This constitutes a greater challenge, especially when reversibility is important. A crystalline sample provides excellent order, but if the coupling between neighbouring molecules is too inflexible then the crystal will crack irreversibly as individual molecules change shape. An amorphous sample is less prone to cracking, but the reduced order makes it more difficult to optimize the arrangement of molecules so that their shape changes are macroscopically cooperative. Nevertheless, several tricks have been identified that impose suitable molecular orientation in amorphous films and liquid-crystalline films to achieve slow macroscopic bending or contraction upon illumination[4].

So, a long-cherished aim of solid-state organic chemistry remains: to discover new single-crystal to single-crystal transformations, with a smooth transition between the crystal lattices of the starting material and

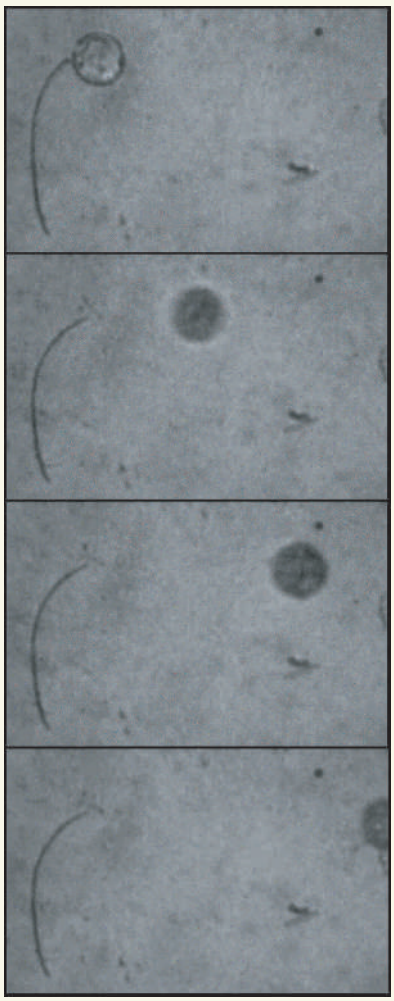


Figure 1 | Light-driven crystal movement. Irie and colleagues[3] have prepared crystals that bend rapidly in response to light. This series of images from the authors' supplemental movie shows how a single crystal fibre, when illuminated by a pulsed laser, bends within 1 millisecond to launch a 50-μm gold ball towards the light source.

the product. For this to happen, the product molecules must form randomly throughout the starting material, rather than establishing larger 'islands' of a separate crystal phase that would greatly disrupt the original bulk lattice. Such phase separation makes cracking all but inevitable, because just a 1% change in lattice spacing, caused by product formation, would force neighbouring lattices within a distance of 50 molecules to fall completely out of register.

There is a handful of special cases in which the shape change that occurs upon reaction is subtle enough to allow individual product molecules to fit into the original lattice. In such compounds, product molecules can be generated at random throughout the starting material by using wavelengths of irradiating light that penetrate the entire crystal[5], rather than using wavelengths that would be strongly absorbed in only a thin surface layer of the compound. The integrity of these crystals is sufficiently preserved to allow the arrangement of atoms in the starting material and the product to be determined by X-rays throughout the course of the transformation. With such crystals it has even been possible to perform several write–erase cycles of optical holograms, involving patterned regions with a higher concentration of product molecules and thus a different refractive index[6].

Irie and co-workers[3] now report rapid, light-induced shape-changing of tiny single crystals. Some of these resemble the rare cases described above — single crystals that change shape but remain intact, even when containing as much as 70% product. But the most impressive examples appear quite different. The authors grew crystal needles (about 200 μm long and 5 μm in diameter) by sublimation so as to be attached at one end to a microscope slide. When illuminated from the side by an ultraviolet laser pulse, the needles bent within a millisecond, displacing the free end by 50 μm. As a dramatic demonstration of this effect, the needles can launch a tiny gold sphere as if it were a tennis ball (Fig. 1; see also the movie in the Supplementary Information for the paper).

The reversibility of this movement is particularly noteworthy — the needles can undergo 80 cycles of photochemical bending and straightening with no apparent damage to the crystal integrity, nor any diminution in displacement amplitude.

To achieve this bending movement, the conversion of starting material to product must occur to different degrees across the needle's width, so that the crystal deforms in the same way as a heated bimetallic strip. As mentioned previously, such localization of conversion along one surface was specifically avoided in previous work, to avoid cracking.

It remains to be explained how organic crystals can survive intact through so many cycles of deformation. In particular, it is difficult to work out which molecular features of this system provide sufficient cohesion between crystal layers to allow bending upon differential reaction, but enough flexibility to avoid cracking. An improved understanding of this unexpected resilience is required, along with a method to assemble microscale components, if we are ever to make a useful chemico-mechanical device from such materials. But simply demonstrating this behaviour suggests a possibility that previously seemed remote. ■

J. Michael McBride is in the Department of Chemistry, Yale University, Box 208107, New Haven, Connecticut 06520, USA.
e-mail: j.mcbride@yale.edu

1. Whittaker, M. *et al. Nature* **378,** 748–751 (1995).
2. Pease, A. R. *et al. Acc. Chem. Res.* **34,** 433–444 (2001).
3. Kobatake, S., Takami, S., Muto, H., Ishikawa, T. & Irie, M. *Nature* **446,** 778–781 (2007).
4. Finkelmann, H., Nishikawa, E., Pereira, G. G. & Warner, M. *Phys. Rev. Lett.* **87,** 15501–15504 (2001).
5. Novak, K., Enkelmann, V., Wegner, G. & Wagener, K. B. *Angew. Chem. Int. Edn Engl.* **32,** 1614–1616 (1993).
6. Köhler, W., Novak, K. & Enkelmann, V. *J. Chem. Phys.* **101,** 10474–10480 (1994).

50 YEARS AGO

"A cæsium clock" — In the issue of *Atoms for Peace Digest* for February 23, the fortnightly periodical of the United States Information Service, brief details are given of the 'Atomichron', the first cæsium atom-beam clock to be available commercially in the United States, and probably in the world... The model costs 50,000 dollars, weighs 500 lb. and measures 84 in. × 22 in. × 18 in. It is capable of keeping time to an accuracy of five seconds over a period of 300 years... The Atomichron has already been used by the U.S. Armed Forces in navigation systems. It should find application in the improvement of astronomical observations, long-range navigation, radio communications, surveying and map-making, and in the study of basic physics, particularly the accurate determination of the velocity of light.
From *Nature* 6 April 1957.

100 YEARS AGO

The following account of a toad attacking a golden carp may be of interest to some of your readers from its bearing on the ancient belief that frogs and toads are at enmity with carp, and kill them by destroying their eyes... On March 29 my son directed my attention to a large golden carp (*C. auratus*) lying in shallow water near the edge of a pond in my garden with a frog or toad apparently resting on its head. The fish appeared to be very sluggish... On examination it was found that the head of the fish was held tightly by a medium-sized common toad (*Bufo vulgaris*), which had obtained a very firm grasp by inserting its fore-limbs as far as the second, or elbow, joint into the sockets of the eyes of the unfortunate fish... A few years ago in the same pond... I found a toad embracing a water-logged puffball so firmly that it required considerable force to release the fungus from the amphibian's grasp.
From *Nature* 4 April 1907.

50 & 100 YEARS AGO

GENOMICS

Global views of leukaemia

Todd R. Golub

Genomic characterization of a type of leukaemia has resulted in the identification of common genetic abnormalities that underlie the disease. The results constitute an advance on several fronts.

It is now widely acknowledged that cancer is a genetic disease, caused largely by the acquisition of mutations in somatic (non-germline) cells after birth — or in some cases during fetal development. A systematic search for these mutations was previously impossible, but progress in genomics technology holds the promise of making the complete characterization of the 'cancer genome' feasible.

An important advance has been the development of high-density DNA microarrays for detecting regions of genomic amplification or deletion. In principle, the application of these arrays might allow for both a global view of recurring abnormalities in genomic copy number, and a sufficiently precise mapping of those abnormalities to allow the gene(s) affected to be identified. On page 758 of this issue, Mullighan and colleagues[1] transform this concept into practice with their discovery of frequent deletions and loss-of-function mutations in the *PAX5* gene in childhood acute lymphoblastic leukaemia (ALL). This disease results from a defect in the differentiation of blood cells: an overproduction of immature B lymphocytes that overtakes the normal development of blood cells in the bone marrow.

By studying 192 ALL samples using DNA microarrays containing probes for about 350,000 genetic loci, the authors were able to rapidly identify deletions of the short arm of chromosome 9 in 57 patients (30%). Because of the high density of probes on these arrays, it was possible to narrow down the target of the deletion to a single gene, *PAX5*, which encodes a transcription factor. Some cases showed tiny deletions affecting only those regions of the gene that code for a particular part of the PAX5 protein, the 'transactivation domain'.

Most cases with a *PAX5* deletion lacked mutations in the remaining copy (allele) of the *PAX5* gene. This is strong evidence for 'haploinsufficiency', the situation in which the single functional allele cannot produce enough of the protein product. The authors hypothesized that some of the ALL cases might suffer from PAX5 loss of function by another mechanism, namely point mutation — the mutation of a single base. Indeed, they discovered *PAX5* point mutations in an additional 14 patients — all but one of the mutations occurred in only one of the two alleles, which again supports the notion of haploinsufficiency. The results with *PAX5* add to a growing number of examples of haploinsufficiency involving transcription factors[2].

Another striking finding[1] was the frequent co-occurrence of *PAX5* loss of function with particular molecular abnormalities — for example, a chromosomal translocation resulting in the fusion of two genes, *ETV6* and *RUNX1* (ref. 3). This is detectable at the time of birth in blood taken from the umbilical cord of infants who go on to develop ALL years later, suggesting that additional genetic hits are required to produce the consequences of *ETV6/RUNX1* fusion[4]. Consistent with this observation, mice genetically engineered to harbour only the *ETV6/RUNX1* defect do not suffer from the hallmarks of leukaemia[5]. So it seems that *PAX5* haploinsufficiency may represent a collaborating event in the development of *ETV6/RUNX1* leukaemias.

Mullighan and colleagues' study has implications for large-scale efforts that aim to characterize the cancer genome, including the Cancer Genome Atlas project sponsored by the US National Institutes of Health. Recent reports have led some to believe either that mutations in cancer genes are exceedingly rare, or that the molecular heterogeneity of cancer will lead to a mountain of genomic data that makes the problem hopelessly complex. But it is evident that when high-density approaches are applied to large numbers of patient samples, clear patterns of recurrent molecular abnormalities emerge.

The study[1] further shows that analyses of genomic copy number, which are relatively inexpensive, can be used to guide the selection of genes for resequencing: until it becomes possible to affordably sequence cancer genomes in their entirety (or at least sequence all protein-coding regions), such prioritization will be necessary. It also highlights the recurrent theme of a connection between cancer and cellular differentiation. Apart from abnormalities in *PAX5*, Mullighan *et al.* found defects in other genes — *EBF, IKZF1, IKZF3, LEF1, TCF3* and *BLNK* (although with much less statistical support) — thought to be involved in regulating lymphocyte differentiation. A systematic

GEOLOGY

Crystal-clear ideas

Caves stumbled upon deep in the workings of the Naica mine in Chihuahua, Mexico, are famed for the huge, elongated crystals of selenite that they contain. The most spectacular such discovery, the transparently named *Cueva de los Cristales*, came in 2000.

The crystals found there — at 300 metres' depth, a temperature of just under 60 °C and 100% humidity — are up to 11 metres long (see picture). Juan Manuel García-Ruiz and colleagues suggest the very specific conditions that were required for such huge crystals to form (J. M. García-Ruiz *et al. Geology* **35,** 327–330; 2007).

Selenite is the colourless crystalline form of calcium sulphate dihydrate ($CaSO_4 \cdot 2H_2O$), better known as gypsum. Gypsum's solubility in water is highest at around 58 °C. Above that temperature, however, it is thermodynamically less stable than the anhydrous form of calcium sulphate, known simply as anhydrite.

According to the authors' ideas, then, the ambient conditions of the *Cueva de los Cristales* would have been ideal for maintaining low-salinity cave-water slightly undersaturated in anhydrite (which would have originated from hydrothermal processes during the formation of the surrounding rock) and slightly supersaturated in gypsum over long periods. This tiny equilibrium imbalance would have encouraged the slow, sparse formation of very long gypsum crystals.

The authors backed up their theory with various hydrochemical and geochemical analyses of samples from the cave. But they point out that the effulgent result of the unique natural circumstances that they describe might not be visible for future generations. The Naica mines today number among the most important for lead and silver in the world. Once their booty is exhausted, however, and water-pumping of the workings stops, nature might reclaim the *Cueva de los Cristales*, reverting it to its erstwhile fully flooded state.

J. TRUEBA/GEOL. SOC. AM.

Richard Webb

search for tumour-associated mutations in developmentally important genes would seem likely to bear fruit.

One might question whether lessons learned from malignancies in blood-cell production will generalize to solid tumours. But there is little evidence, if any, that blood malignancies are in any way unique. For example, chromosomal translocations, which were initially thought to be a curiosity of acute leukaemias, are now known to be common in prostate cancer. Mutations that activate protein-kinase enzymes, and that frequently cause them to become oncoproteins, were again initially associated with leukaemia, but are now commonly recognized in solid tumours. And cancer stem cells, now detectable in breast cancers, and in brain and other tumours, were at first seen as relevant only to leukaemia. Cancer pathogenesis seems to follow a largely common path, regardless of the type of cell in which it originates.

Finally, we are left with an obvious challenge: devising strategies for fixing the defects in cellular differentiation caused by recurrent loss-of-function mutations. Treatment with all-*trans* retinoic acid is an effective 'differentiation therapy' for acute promyelocytic leukaemia, although this has been generally regarded as relevant only to that subtype of leukaemia. But if defects in differentiation underlie many (or possibly all) cancers, a broader exploration of such treatments should perhaps be entertained. Whereas a clear path has been established for developing anticancer drugs that inhibit the increased activity of oncoproteins such as kinases, approaches to restoring lost function in tumour-suppressor proteins are less obvious. And the pharmacological modulation of transcription factors is in large measure uncharted territory.

Innovative approaches to chemical screening and drug discovery are likely to be needed. But the fact that *PAX5* abnormalities in ALL seem to exist mainly in the haploinsufficient state may make the problem easier — finding small molecules that augment the expression or activity of the residual *PAX5* allele is probably feasible.

In any event, we have a challenge that we should be pleased to tackle. Trying to develop better therapies for cancer without knowing the underlying genetic causes would be utterly impossible. ■

Todd R. Golub is at the Broad Institute of Harvard and the Massachusetts Institute of Technology; and the Dana-Farber Cancer Institute, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA.
e-mail: golub@broad.harvard.edu

1. Mullighan, C. G. *et al. Nature* **446,** 758–764 (2007).
2. Song, W. J. *et al. Nature Genet.* **23,** 166–175 (1999).
3. Golub, T. R. *et al. Proc. Natl Acad. Sci. USA* **92,** 4917–4921 (1995).
4. Ford, A. M. *et al. Proc. Natl Acad. Sci. USA* **95,** 4584–4588 (1998).
5. Tsuzuki, S., Seto, M., Greaves, M. & Enver, T. *Proc. Natl Acad. Sci. USA* **101,** 8443–8448 (2004).

BIOPHYSICS

Quantum path to photosynthesis

Roseanne J. Sension

Knowing how plants and bacteria harvest light for photosynthesis so efficiently could provide a clean solution to mankind's energy requirements. The secret, it seems, may be the coherent application of quantum principles.

The Sun bombards Earth with a steady stream of energy — about 10^{17} joules per second, on average[1]. If we could convert this energy efficiently to a chemically useful form, our reliance on fossil fuels and nuclear power — and so our production of climate-change agents and hazardous waste materials — could be substantially reduced. But how can we achieve such efficiency?

Of course, the photosynthesizing organisms on Earth already have the answer. In higher plants and certain bacterial systems, the initial steps of natural photosynthesis harness light energy with an efficiency of 95% or more — values that we can only aspire to with artificial photocells. Elsewhere in this issue, Engel *et al.*[2] (page 782) take a close look at how nature, in the form of the green sulphur bacterium *Chlorobium tepidum*, manages to transfer and trap light's energy so effectively. The key might be a clever quantum computation built into the photosynthetic algorithm.

Photosynthesis is initiated by the excitation, through incident light, of electrons in pigment molecules — chromophores — such as chlorophyll. This electronic excitation moves downhill from energy level to energy level through the chromophores before being trapped in a reaction centre, where its remaining energy is used to initiate the production of energy-rich carbohydrates. In natural light-harvesting systems, chlorophyll pigments are arranged together in an 'antenna', sometimes with elegant symmetry and sometimes with apparent randomness, but always with a precise structure that is supplied by a protein scaffold (Fig. 1). Many studies indicate[3–6] that this strictly defined structure, together with the close proximity of the chromophores, is essential for the strong absorption of light, fast

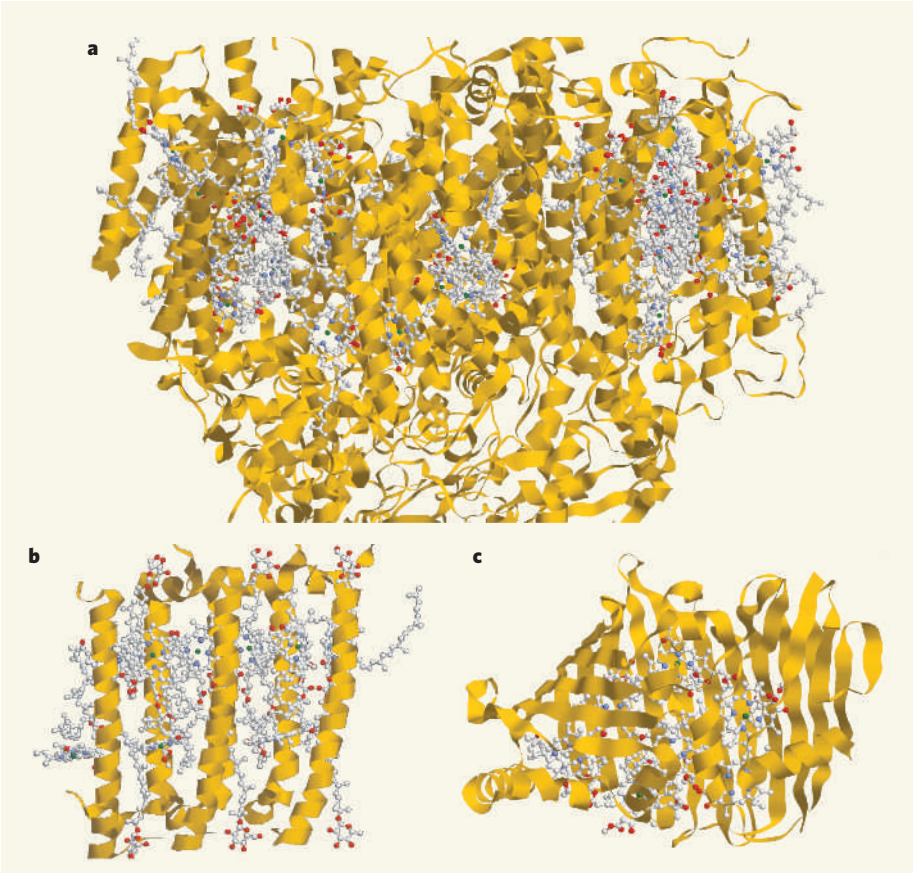


Figure 1 | Complex complexes. Three representative pigment protein complexes involved in natural photosynthesis. **a,** Photosystem II, found in oxygenic photosynthetic organisms[8]. **b,** Light-harvesting complex 2, found in purple bacteria[9]. **c,** The Fenna–Matthews–Olsen antenna complex[10], the bacteriochlorophyll complex investigated by Engel *et al.*[2]. Yellow, protein; white, chromophores.

downhill energy transfer and efficient trapping of the electron excitation in a reaction centre, all of which are characteristic of natural photosynthesis.

But where does quantum mechanics, let alone quantum computing, fit in here? The mechanism of energy transfer through chromophore complexes has generally been assumed to involve incoherent hopping — that is, seemingly uncoordinated movement in a 'random walk' with a general downhill direction — either between individual chromophores, or between modestly delocalized energy states spanning several chromophores. The energy transfer is determined by quantum-mechanical states and their overlaps, to be sure, but there is nothing inherently 'quantum' or wave-like in the process itself.

Engel *et al.*[2], however, performed two-dimensional Fourier transform spectroscopy of the bacteriochlorophyll Fenna–Matthews–Olsen antenna complex, and discovered regular variations in the intensity of their signal. These 'quantum beats', which persist for hundreds of femtoseconds, are characteristic of coherent coupling between different electronic states. In other words, the electronic excitation that transfers the energy downhill does not simply hop incoherently from state to state, but samples two or more states simultaneously. The data also suggest that the protein scaffold might itself be structured to dampen fluctuations that would induce decoherence of the electronic excitation.

Coherent energy transfer allows the 'wave-like' sampling of the energy landscape to establish the easiest route for the electronic excitation to the reaction complex much faster than the semi-classical hopping mechanism allows — indeed, it does so almost instantaneously. The process is analogous to Grover's algorithm in quantum computing, which has been proved to provide the fastest possible search of an unsorted information database[7].

Although the data were acquired at low temperature (77 kelvin), the observation of electronic coherences in such a complex system is remarkable. Assuming that the effect is general — that similar coherences occur in many different natural light-harvesting systems, and are observed at non-cryogenic temperatures — we may find that nature, through its evolutionary algorithm, has settled on an inherently quantum-mechanical process for the critical mechanism of efficient light harvesting. This is an interesting lesson to be considered when designing artificial systems for this purpose. ■

Roseanne J. Sension is in the FOCUS Center and Department of Chemistry, University of Michigan, 930 North University Avenue, Ann Arbor, Michigan 48109-1055, USA.
e-mail: rsension@umich.edu

1. Grätzel, M. *Chem. Lett.* **34,** 8–13 (2005).
2. Engel, G. S. *et al. Nature* **446,** 782–786 (2007).
3. van Amerongen, H., Valkunas, V. & van Grondelle, R. *Photosynthetic Excitons* (World Scientific, Singapore, 2000).
4. Yoder, L. M., Cole, A. G. & Sension, R. J. *Photosynth. Res.* **72,** 147–158 (2002).
5. Shiang, J. J., Yoder, L. M. & Sension, R. J. *J. Phys. Chem. B* **107,** 2162–2169 (2003).
6. Brixner, T. *et al. Nature* **434,** 625–628 (2005).
7. Grover, L. K. *Phys. Rev. Lett.* **79,** 325–328 (1997).
8. Ferreira, K. N., Iverson, T. M., Maghlaoui, K., Barber, J. & Iwata, S. *Science* **303,** 1831–1838 (2004).
9. Papiz, M. Z., Prince, S. M., Howard, T., Cogdell, R. J. & Isaacs, N. W. *J. Mol. Biol.* **326,** 1523–1538 (2003).
10. Camara-Artigas, A., Blankenship, R. & Allen, J. P. *Photosynth. Res.* **75,** 49–55 (2003).

CELL BIOLOGY

Fraternal twins

Franck Duong

A popular route for protein transport into and across cell membranes is through the Sec channel. This channel seems to function by forming a dimer of two identical units, where each has a distinct role.

The transmembrane Sec channel, or Sec translocon, is a major protein-transport route across the endoplasmic reticulum of higher organisms and the cell membrane of bacteria. This essential machinery ensures the correct distribution of cellular proteins, and catalyses the translocation, and membrane integration, of hundreds of different proteins that carry a specific targeting signal called the signal sequence. To mediate transport, the Sec channel associates with different partners in the cell's internal fluid, or cytosol, that supply the driving force for translocation. For example, during protein translation, the ribosome — the factory for all protein production — feeds nascent polypeptide chains directly into this channel. In bacteria, the channel also associates with an enzyme called SecA (or SecA ATPase), which, following translation, 'pushes' the protein substrate into the channel. A report by Osborne and Rapoport[1], published in *Cell*, provides a view of how SecA and the Sec channel work together.

The fact that the Sec channel forms both a membrane conduit for polypeptides and a binding site for its translocation partners has been rationalized through structural analysis of its evolutionarily conserved core component — the heterotrimeric SecY–SecE–SecG complex, called SecY for short[2]. The atomic structure of the Sec channel revealed four domains: an hourglass-shaped, hydrophilic conduit located in the body of the SecY complex; a constricted 'pore ring', which seals the conduit in the middle; a 'plug' domain that lies on top of the constricted region when the channel is inactive; and a 'lateral gate' serving as both the binding site for the signal sequence and an escape route towards the lipid layer for transmembrane protein domains (Fig. 1, overleaf). On the part of the complex facing the cytosol, at least two large loops that extend out of the plane of the membrane serve as a docking site for either SecA or ribosomes. Biochemical analyses using a translocation system reconstituted *in vitro* have provided experimental support for some of the proposed functions of these structural domains[3–5].

From such structural information, one would predict that a single SecY copy would be enough to perform the transport task. But that possibility had already been ruled out by numerous observations of membrane-embedded and solubilized SecY complexes that naturally form oligomers containing two (dimer) or four (tetramer) copies of the complex. Although there now seems to be a consensus that the bacterial Sec channel exists as a dimer, and that its mammalian counterpart is tetrameric when interacting with ribosomes[6,7], the underlying reason for the oligomerization of SecY complexes had remained a puzzle. Experimentally, it is difficult to monitor the oligomeric state of membrane proteins and the function of their different states. But Osborne and Rapoport[1] have now provided substantial insights into this problem.

By cross-linking the polypeptide substrate to cysteine amino-acid residues strategically engineered and positioned in the complex, the authors first demonstrated that the lateral gate and the plug domain of only one SecY complex simultaneously make contact with the translocating polypeptide chain and its signal sequence. This confirms the structural prediction that a single SecY copy in the SecY translocon forms the protein-conducting channel.

Next, the authors reasoned that, if two SecY copies work together but each has a distinct function, an inactive SecY complex should be rescued on interaction with its active counterpart. This result would indeed prove that the oligomers are involved in translocation. To facilitate such analysis, two *secY* genes have previously been fused in tandem, resulting in the production of a covalently linked SecY dimer[8]. Using this approach, Osborne and Rapoport showed that an inactivating mutation in one SecY copy does not prevent protein translocation as long as a normal copy of this complex is also present in the tandem construct.

It remains conceivable that one SecY copy forms an active unit by interacting with another SecY copy, but without the contribution

ECOLOGY

Poole resources

M. LANE/ALAMY

The Manila clam has been on the move for decades. It is native to the western Pacific, but following introduction to other parts of the Pacific, and then to southern Europe, it was brought to Britain in the 1980s as a source of seafood. At one site in Britain, Poole Harbour in Dorset, the clam (*Tapes philippinarum*) has now become naturalized.

This could be worrying: when colonizing fresh regions, invasive species may devastate components of the existing flora and fauna. For the Eurasian oystercatcher, however, the advent of the clam at Poole is good news, as Richard Caldow and colleagues report (R. G. W. Caldow *et al. Proc. R. Soc. B* doi: 10.1098/rspb.2007.0072). This species of bird, *Haematopus ostralegus ostralegus* (pictured), overwinters in Poole Harbour. From their observations of its feeding habits, and from modelling studies, the authors conclude that the oystercatchers have benefited considerably from the extra source of food.

Their observations show that a large proportion of the overwintering oystercatcher population of around 1,200 feeds on the clams, a habit not previously recorded, and that clam meat constitutes a notable part of the birds' diet.

The simulations were carried out with an 'individuals-based' model of shorebird foraging, with the aim of providing a population-level estimate of the effect of the additional food source. The predicted result is a significant reduction in the mortality of the birds, which face the prospect of starvation in the period between September and March.

As yet, there is no evidence that the Manila clam has affected other species of bivalve at Poole, although it occurs at low densities there compared with populations of the species elsewhere. But the clams' occupation of this northern site was probably made possible by locally warm sea temperatures, and Caldow *et al.* raise the inevitable question of what consequences a continued warming might have. They envisage a further spread north. That process might exacerbate the retreat of the cold-water species that currently constitute food sources for shorebirds. But if, like the Eurasian oystercatcher, other birds develop a taste for the clam, the results might even be beneficial — at least from the avian point of view.

Tim Lincoln

of the other half of the fusion protein. However, the results provide clear evidence that two SecY complexes are necessary to form an active channel. Moreover, because the inactivating mutation was previously shown to affect the binding of SecA to the channel[9], such complementary action could be explained only if SecA binds to one SecY copy and the protein substrate crosses the membrane through the other. Osborne and Rapoport confirm such a structural asymmetry. In the tandem construct, one SecY copy is cross-linked with the translocating protein substrate, and the other copy is cross-linked near the ATPase motor domain of SecA.

On the basis of these findings, Osborne and Rapoport[1] provide a refined model of SecA-mediated protein translocation (Fig. 1). One SecY complex serves as the protein-conducting channel, whereas its non-translocating counterpart forms a static docking site for the ATPase motor domain of SecA. This model is consistent with results[10] showing that a single SecY complex is sufficient to bind to SecA. Taking into consideration the dimensions of the SecY dimer and SecA, the authors propose that the deep groove observed in the crystal structure of SecA (ref. 11), and postulated to be involved in binding to signal sequences and polypeptide chains, would be located just below the active copy of SecY in the channel. This would be an optimal position for pushing the protein substrate through.

How exactly SecA, which also forms a dimer in solution, binds to the Sec channel, and how it converts chemical energy into mechanical work, remains to be discovered. But the present study[1] is a milestone on the way to understanding the intricate organization of the translocon. It reveals once again the unique characteristics of this remarkable machine. ■

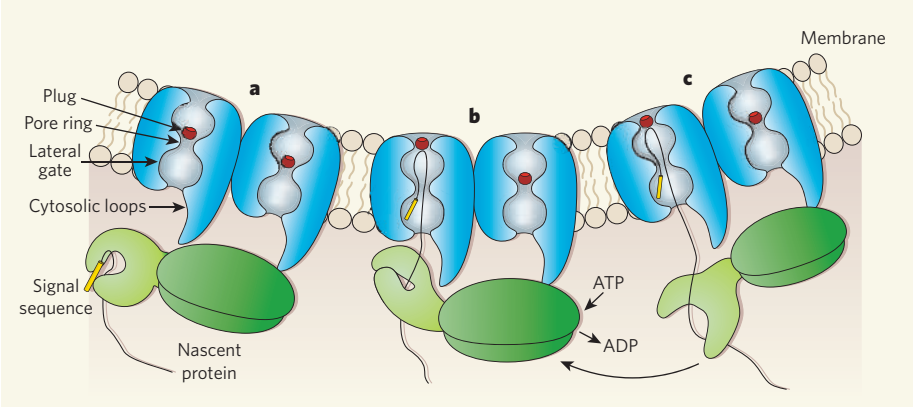


Figure 1 | The protein translocation process as proposed by Osborne and Rapoport[1]. This simplified representation shows some of the structural elements of the dimeric SecY channel (blue) and the SecA ATPase (green), from a cut-away view of the membrane. **a,** At the beginning of the reaction, the conduits of the SecY dimer are sealed by the constricted pore ring and the plug domain (red). The protein substrate and its signal sequence (yellow) are engaged with the protein-binding domain of SecA (light green), whereas its ATPase domain (dark green) is anchored to the cytosolic loops of one SecY copy. **b,** On binding and hydrolysis of ATP, SecA pushes the signal sequence as a hairpin loop into the neighbouring SecY copy. The insertion of the hairpin causes the plug to move away from the centre of the conduit and fixes the pore in the open state. **c,** The ATPase domain of SecA remains anchored to one copy of SecY, and its protein-binding domain grasps another, downstream, segment of the polypeptide chain. The ATP-dependent cycle is repeated until the polypeptide is entirely transferred across the membrane.

Franck Duong is in the Department of Biochemistry and Molecular Biology, Life Sciences Institute, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada.
e-mail: fduong@interchange.ubc.ca

1. Osborne, A. R. & Rapoport, T. A. *Cell* **129,** 97–110 (2007).
2. Van den Berg, B. *et al. Nature* **427,** 36–44 (2004).
3. Plath, K. *et al. Cell* **94,** 795–807 (1998).
4. Cannon, K. S. *et al. J. Cell Biol.* **169,** 219–225 (2005).
5. Tam, P. C., Maillard, A. P., Chan, K. Y. & Duong, F. *EMBO J.* **24,** 3380–3388 (2005).
6. Mitra, K. *et al. Nature* **438,** 318–324 (2005).
7. Ménétret, J. F. *et al. J. Mol. Biol.* **348,** 445–457 (2005).
8. Duong, F. *EMBO J.* **22,** 4375–4384 (2003).
9. Mori, H. & Ito, K. *Proc. Natl Acad. Sci. USA* **98,** 5128–5133 (2001).
10. Alami, M., Dalal, K., Lelj-Garolla, B., Sligar, S. G. & Duong, F. *EMBO J.* doi:10.1038/sj.emboj.7601661 (2007).
11. Osborne, A. R., Clemons. W. M. Jr & Rapoport, T. A. *Proc. Natl Acad. Sci. USA* **101,** 10937–10942 (2004).

OBITUARY

Knut Schmidt-Nielsen (1915–2007)

Quests in comparative physiology.

Knut Schmidt-Nielsen, one of the all-time greats of animal physiology, died on 25 January. His influence on the subject was profound, not only through his original research but also through his books and the meetings he organized.

Schmidt-Nielsen was born in Norway but moved to the United States in 1946. Much of his finest work concerns life in deserts, where the combination of heat and drought makes survival particularly difficult. The most effective means of keeping cool depend on evaporation of precious water, either in the breath or as sweat, but kangaroo rats survive in the Arizona Desert with nothing at all to drink. They keep reasonably cool by spending the day in burrows and emerging only at night. But even at night they would lose much too much water in their breath if it were not for their remarkable noses. The air they breathe in is relatively cool and dry, but it is inevitably warmed in the body and becomes saturated with water vapour. To minimize water loss, this air must be cooled before it leaves the body to condense out most of the vapour.

With his first wife Bodil, Schmidt-Nielsen showed that the incoming air cools the surfaces of the nasal cavity, which, in turn, cool the air when it is breathed out again. He showed that the same principle operates in other mammals and birds, but is particularly effective in kangaroo rats because their nasal surfaces are enlarged by elaborate nasal bones known as turbinals. They also save water by producing exceptionally concentrated urine.

The countercurrent effect in the nose is vital to desert rodents. But it is potentially troublesome to dogs overheated by exercise, as they need to let water evaporate to cool themselves. With colleagues, Schmidt-Nielsen showed that panting dogs avoid the effect by breathing in through the nose but out through the mouth.

Unlike kangaroo rats, camels are too big to retire to burrows in the heat of the day. Working in the Sahara Desert, the Schmidt-Nielsens showed that camels avoid the water loss that evaporative cooling would incur by allowing their bodies to heat up by day and cool by night. A camel may start the morning with a body temperature of only 34 °C, but warms to 41 °C during the afternoon. This strategy would be ineffective for small animals such as kangaroo rats, because they would quickly heat up to lethal temperatures, but it works well for camels. Later, in Australia, Schmidt-Nielsen showed that camels' noses conserve water even more effectively than do those of kangaroo rats; hygroscopic surfaces dry the outgoing air.

Sea birds face a different problem: they have nothing to drink but sea water, which has a much higher osmotic concentration than their blood. In principle, they could produce urine even more concentrated than sea water, but their kidneys are not equal to the task. Instead, Schmidt-Nielsen found that glands opening into their nostrils secrete droplets of concentrated salt solution that are shed by a shake of the head.

With William Bretz, he tackled another aspect of bird physiology — the operation of their lungs. The lungs of birds are very different from those of mammals: instead of being a mass of blind-ended sacs they are formed from slender parallel tubes. Mammalian breathing is a reciprocating process; the air leaves by the same route that it came in. However, Schmidt-Nielsen and Bretz showed that airflow is one-way in bird lungs, flowing in at one end of the tubes and out at the other, and they suggested that gas exchange is enhanced by a countercurrent flow of air and blood. This would help birds to get oxygen into the blood at the high rates needed to power active flight.

In the later part of his career, Schmidt-Nielsen concentrated on the consequences for animals of differences in body size. With C. R. (Dick) Taylor and other colleagues, he measured the rates of oxygen consumption of various mammals, ranging from mice to horses, running at different speeds. They calculated the quantity of oxygen used for each kilogram of animal moving a distance of one metre, and found that this was less for larger animals, falling off in proportion to $(\text{body mass})^{-0.4}$.

Data for flying birds and swimming fish showed similar relationships to body mass, although flight is cheaper than running (for the same body mass), and swimming is cheaper still. These findings, and later work on this theme by others, have given valuable insight into the energy budgets of animals. They have also caused puzzlement (still not wholly dissipated), because they imply that the muscles of small animals work less efficiently than those of large ones.

Of his books, *Desert Animals* (1964) and *Scaling* (1984) cover Schmidt-Nielsen's two main fields of interest. His *Animal Physiology* (1960 and later editions) has been used and appreciated by countless students. *How Animals Work* (1972), my favourite among his books, is a small masterpiece. *The Camel's Nose* (1998) is his autobiography.

DUKE UNIV. PHOTOGRAPHY

Schmidt-Nielsen understood the value of bringing scientists together. Like the other participants, I benefited enormously from the series of select meetings that he and Liana Bolis arranged at delightful locations in Europe. He convened the small preliminary meeting that laid the foundation for the 1975 'Scaling Conference' in Cambridge, UK, a meeting that inspired much subsequent research. He was president of the International Union of Physiological Sciences from 1980 to 1986.

A fine bronze statue celebrating Schmidt-Nielsen's work with desert animals in particular, which is shown in the picture here, stands outside the Department of Biology at Duke University in Durham, North Carolina, where he had worked since 1952. He received many other honours, including the International Prize for Biology, presented at an impressive ceremony by the Emperor of Japan. He was nevertheless unassuming, with the slight diffidence of manner characteristic of the old-fashioned ideal of the perfect gentleman. He and his wife Margareta were among the most delightful hosts I have known. ■

R. McNeill Alexander
R. McNeill Alexander is at the Institute for Integrative and Comparative Biology, University of Leeds, Leeds LS2 9JT, UK.
e-mail: r.m.alexander@leeds.ac.uk

Q&A

CELL BIOLOGY

Autophagy and cancer

Beth Levine

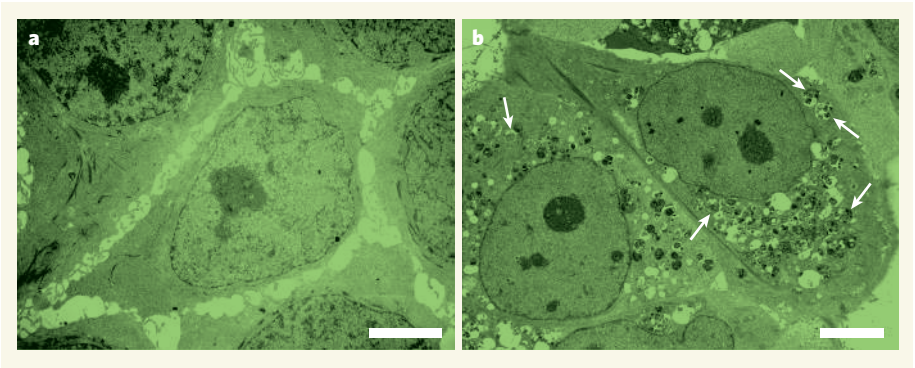
Autophagy is the degradation of redundant or faulty cell components. It occurs as part of a cell's everyday activities and as a response to stressful stimuli, such as starvation. Connections with cellular life-and-death decisions and with cancer are now emerging.

How is autophagy induced?

Autophagy occurs when cells need to 'self-cannibalize' or degrade their constituents. Underlying 'housekeeping' levels of autophagy probably occur in most normal cells to prevent the accumulation of protein aggregates and defective cellular substructures. Certain environmental cues (such as starvation, high temperature, low oxygen, hormonal stimulation) or intracellular stress (damaged organelles, accumulation of mutant proteins, microbial invasion) activate signalling pathways that increase autophagy. Classically, most research on how autophagy is induced has focused on an enzyme called TOR kinase. This enzyme is a sensor of nutrient status and a master regulator of cell growth; it negatively regulates autophagy through its effects on a set of proteins known as autophagy-execution proteins. However, it is now clear that numerous signalling pathways, such as those involved in the control of cell growth, DNA-damage repair, a form of programmed cell death called apoptosis, and immunity, can also induce autophagy. It is still a mystery how these specific signals turn on the autophagic machinery.

What happens once autophagy is induced?

Once the cell receives the appropriate signal, the autophagy-execution proteins trigger a cascade of reactions that result in membrane rearrangements to form a double-membrane-bound vesicle called an autophagosome (Fig. 1). Initially, an 'isolation membrane' forms, although its origin is still controversial. The membrane surrounds the cytoplasmic contents to be degraded, and its edges fuse to form the autophagosome. This vesicle then fuses with a lysosome (or a vacuole in yeast), with the release of lysosomal digestive enzymes into the lumen of the resulting autolysosome. The sequestered cytoplasmic contents are degraded inside the autolysosome into free nucleotides, amino acids and fatty acids, which are reused by the cell to maintain macromolecular synthesis and to fuel energy production. The nutrient recycling and housekeeping functions of autophagy promote cell survival, although in certain contexts autophagy may also promote cell death (Fig. 2, overleaf).



B. LEVINE

Figure 1 | Autophagy in action. **a**, Untreated breast cancer cell (lacking autophagosomes); **b**, breast cancer cell treated with the drug tamoxifen, and containing numerous autophagosomes (arrowed). Scale bars, 5 μm.

Does autophagy also stop protein synthesis?

No. On the contrary, one of its evolutionarily conserved functions, through protein recycling, is to help maintain the synthesis of essential proteins when external nutrients are limited. Although certain stress stimuli that induce autophagy, such as starvation, turn off general protein synthesis, they also turn on the synthesis of specific stress-response proteins, including autophagy-execution proteins. So in this setting, the cell uses a coordinated strategy. To ensure that it has enough amino acids to synthesize the proteins that are essential for its survival, general protein translation is shut down and autophagy is activated.

What are autophagy-specific genes?

These are genes that are required for the execution of the autophagy pathway. Several of them encode proteins that are components of kinase complexes, which regulate the activity of proteins and lipids through the addition of a phosphate group. Alternatively, they encode components of protein-conjugation systems, which attach to each other or to membrane lipids to form the membrane of the autophagosome. Deletion of an autophagy-specific gene blocks autophagy in a cell or organism. These genes were first identified through genetic screens in yeast that included a search for genes that are essential for survival during starvation. Many of these yeast genes are also present in higher organisms, as are the underlying molecular mechanisms of autophagy. Although there seems to be a universal requirement in autophagy for 'autophagy-specific' genes, this does not mean that these genes are not involved in other cellular processes.

What is the difference between autophagy and apoptosis?

These two processes have long been classified as different forms of programmed cell death. But whereas apoptosis invariably leads to cell death, autophagy (despite its frequent occurrence in dying cells) commonly contributes to cell survival. Deletion of autophagy-specific genes in cells of diverse organisms increases cell death during development, as well as susceptibility to starvation and other apoptotic stimuli. This is not to say that autophagy cannot also be a death programme. However, it seems to take on this function primarily when the cellular apoptotic machinery is crippled, or when autophagy is induced to such high levels that cells literally eat themselves to death.

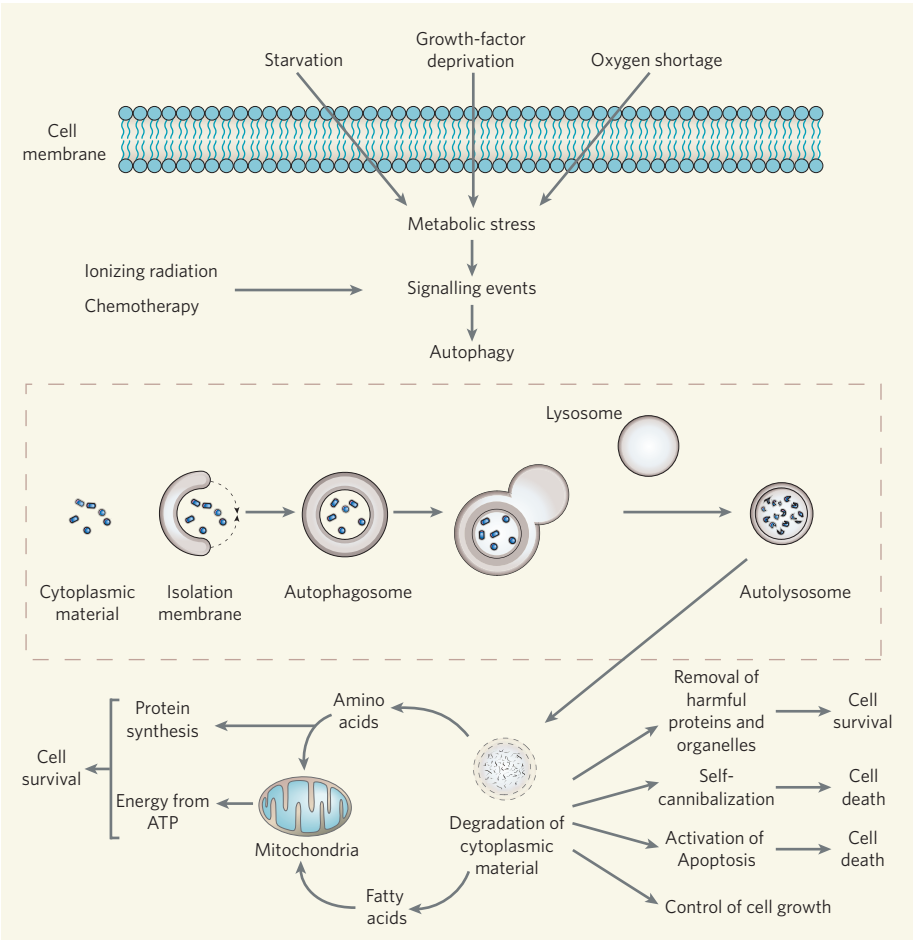


Figure 2 | The autophagy pathway and its diverse cellular functions. Both metabolic stress and cancer therapies activate signalling pathways that stimulate autophagy. The process involves the sequestration of cytoplasmic material by a membrane-bound vesicle called the autophagosome, which then fuses with a lysosome to form an autolysosome. The degradation of cytoplasmic material within the autolysosome can promote cell survival either by generating free fatty acids and amino acids, which can be reused by the cell to maintain energy production and protein synthesis, or by removing harmful proteins and organelles. It can also promote cell death independently (presumably through self-cannibalization) or together with apoptosis. Furthermore, the turnover of proteins and organelles by autophagy may contribute to the control of cell growth.

These scenarios may not be relevant to normal development or to physiological adaptations of cells and tissues to stress. However, they may be relevant to the development of cancer and to cancer therapy, because cancer cells often have mutations that confer resistance to apoptosis, and many chemotherapeutic agents that are toxic to the cell induce high levels of autophagy.

Are autophagy and apoptosis interlinked?

There is a complex, and not fully understood, relationship between them that may vary depending on the biological context. The two pathways are regulated by common factors; they share common components; they can exert overlapping functions; and one pathway may regulate and modify the activity of the other. Many signals that have long been known to activate apoptosis (sphingolipids, death-receptor signalling molecules, serine/threonine death kinases and mitochondrial-associated cell-death proteins) are now known to activate autophagy. Conversely, signalling pathways that inhibit apoptosis (the class I PI3K/Akt signalling pathway and the stress-activated NF-κB signalling pathway) also inhibit autophagy. Intriguingly, regulators of apoptosis, for example members of the anti-apoptotic Bcl-2 family, can also directly 'disarm' autophagy-execution proteins such as Beclin 1. And there is evidence that at least one 'autophagy-specific' protein (Atg5), when cleaved, can activate an apoptotic programme. Beclin 1 also has a structural domain that is considered a hallmark of pro-apoptotic proteins.

Are autophagy and apoptosis mutually exclusive?

No — they commonly occur in the same cell, both when autophagy is trying to keep cells alive and when it contributes to cell death. This co-occurrence can have different consequences. In some circumstances, including starvation and treatment with certain DNA-damaging agents, autophagy delays the onset of apoptosis. In others, such as HIV infection, autophagy is required for the onset of apoptosis in uninfected lymphocytes. During embryonic development in mammals, autophagy neither delays nor promotes apoptosis, but it is required for dying cells to generate signals that ensure the efficient removal of apoptotic corpses. The complex interplay between autophagy and apoptosis is only now beginning to be unravelled.

Is autophagy linked to disease?

Yes, it is linked to both health and disease. Normally, it contributes to adaptation to cellular stress, to development and differentiation, to immunity and to longevity. Too much or too little autophagy can contribute to certain cardiac- and skeletal-muscle diseases, liver diseases, infectious diseases, neurodegenerative disorders and cancer.

What is the cancer connection?

There are really two connections, one at the level of cancer development and the other at the level of cancer treatment. Inactivation of autophagy-specific genes, such as *beclin 1*, results in increased tumorigenesis in mice, and enforced expression of such genes (*beclin 1*, *atg5*) inhibits the formation of human breast tumours in mouse models. Furthermore, net deletions of several autophagy-specific genes are commonly found in human malignancies. Thus, autophagy may be a tumour-suppressor pathway, and its decreased activity may contribute to the development of human cancer. Consistent with this theory, tumour-suppressor genes that are frequently mutated in human cancer (*p53*, *PTEN*) turn autophagy on, and genes that are frequently activated in cancer, such as those encoding class I PI3K and AKT, turn it off. The conflicting pro-survival and pro-death functions of autophagy make the connection to cancer treatment more complex. The pro-survival function may help cancer cells to survive in nutrient-limited environments, and to resist ionizing radiation and chemotherapies. But the pro-death function may help to kill cancer cells, either spontaneously or when they are exposed to radiation or chemotherapy. These seemingly paradoxical functions have fuelled intense debate as to whether autophagy is cancer's friend or foe.

How can autophagy prevent cancer?

There are several possible ways (Fig. 3). The most obvious one is by suppressing tumorigenesis through its death-promoting effects. However, it is unclear whether, like apoptosis, autophagy is an important cell-death pathway in tissue maintenance, because autophagy-deficient organisms often show increased, rather than decreased, cell death. Another hypothesis is that autophagy prevents DNA damage, presumably through its cellular housekeeping role in removing sources of oxidative stress such as defective cellular organelles (mitochondria or endoplasmic reticulum). This theory could mechanistically

link the tumour-suppressor and anti-ageing functions of autophagy and help to explain the increased incidence of cancer as people age and their underlying housekeeping levels of autophagy decline. A third possibility is that it negatively regulates cell growth. Thus, not only might alterations in protein degradation, which is mediated by a complex known as the proteasome, contribute to the development of cancer, but imbalances between autophagy-dependent protein degradation and synthesis might also be a factor. Interestingly, mice with a mutation in one copy of an essential autophagy gene show no defects in apoptosis, but do show increased cellular proliferation in tissues that are prone to tumour formation. This suggests that the effects of autophagy on cell growth rather than its potential pro-death effects may play a pivotal role in tumour suppression.

How can autophagy promote cancer?

Again, there are several possible ways (Fig. 3). In all eukaryotic organisms, autophagy-specific genes promote the survival of normal cells during nutrient starvation. Similarly, autophagy might enhance the survival of rapidly growing cancer cells that have outgrown their vascular supply and face oxygen shortage or metabolic stress. In other words, it may be the crutch that tumour cells use to survive the shortage of energy and nutrients. It might also promote the survival of cancer cells by targeting damaged mitochondria and other organelles for lysosomal degradation, thereby buffering oxidative stress that can be triggered by activated cancer-causing genes or by cancer treatments.

Would you turn autophagy on or off to kill a tumour cell?

Paradoxically, either action may be correct (Fig. 4). The absence of autophagy increases susceptibility to death when cells confront stressful conditions (metabolic stress, cytotoxic chemotherapy). By contrast, high levels of autophagy, or its occurrence in cells with crippled apoptotic machinery, can also lead to cell death. Yet it may be premature to conclude

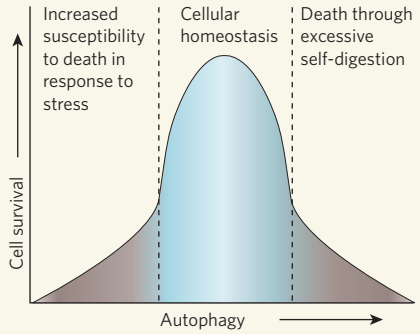


Figure 4 | Relationship between the levels of autophagy and cell death. Physiological levels of autophagy are essential for normal cellular homeostasis. The absence of autophagy increases cell death during metabolic stress and on treatment with cytotoxic chemotherapeutic agents. By contrast, excessive levels of autophagy promote cell death, presumably via self-cannibalization. Antitumour effects may be observed at all levels of autophagy, in relation either to cell death (when autophagy is very low or very high) or to cell-death-independent tumour-suppressor effects (when autophagy levels are intermediate).

that we should try to kill tumour cells by turning autophagy either on or off. Regardless of the approach taken, it is potentially risky if cell death does not invariably occur. Actions that reduce only some of autophagy's effects could result in loss of its tumour-suppressive function, whereas interventions that enhance only some of its effects could result in the increased survival of tumour cells.

Can targeting autophagy be exploited in cancer treatment?

Yes, but only if we can figure out whether autophagy should be turned on or off. Given the numerous paradoxes related to its role in cell death and cancer, this may prove quite difficult. Animal studies can help decipher whether the pro- or anticancer effects of autophagy prevail in any given setting. For example, in one animal model in which tumour cells have a defect in the apoptotic pathway, autophagy promotes tumour-cell survival during metabolic stress, but also prevents tumorigenesis, necrosis and inflammation. Thus, at least in this case, the tumour-suppressor effects of autophagy overshadow the autophagy-dependent survival pathway. By contrast, in another animal model in which tumours are caused by an activated cancer-causing gene, autophagy inhibition with the drug chloroquine enhances therapy-induced apoptosis and tumour regression. These examples illustrate the complexity of predicting the net effect of autophagy in cancer. They also hint at some of the variables that might be involved, including the stage of tumorigenesis, the sensitivity of the tumour cells to apoptosis and the specific molecular alterations in the cells.

What are the next steps in research?

They will be to determine more precisely *if*, *when* and *how* autophagy prevents or promotes cancer.

Will clinical trials help resolve the autophagy and cancer controversy?

Ultimately, the question of whether to turn autophagy on or off may be resolved by information from clinical trials in cancer patients. Many anticancer agents are potent inducers of autophagy. The optimist may view this as evidence that inducing autophagy is a desirable target for cancer therapy. By contrast, the pessimist can rightly argue that autophagy may function as a stress response to counter the toxic effects of antitumour drugs, and that such drugs might work even better if coupled with autophagy inhibitors. There is experimental support for each of these views. But because all current antitumour agents target pathways other than autophagy, it is difficult to dissect the role of autophagy stimulation in their therapeutic action. New agents that specifically target autophagy will have to be developed and tested to resolve this controversy. Only then will we begin to understand how to translate advances in our knowledge of the autophagy pathway into approaches to combat cancer. ■

Beth Levine is in the Department of Internal Medicine, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390-9113, USA.
e-mail: beth.levine@utsouthwestern.edu

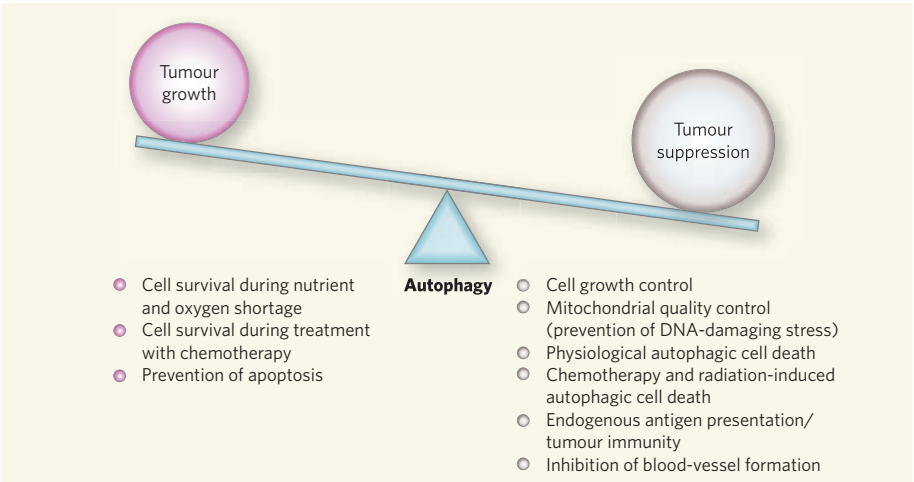


Figure 3 | Conflicting effects of autophagy on cancer. Experimental evidence supports a role for autophagy in both cancer development and suppression. What tips the balance is not clear.

FURTHER READING
Gozuacik, D. & Kimchi, A. *Oncogene* **23,** 2891–2906 (2004).
Hippert, M., O'Toole, P. S. & Thorburn, A. *Cancer Res.* **66,** 9349–9351 (2006).
Levine, B. *Autophagy* **2,** 65–66 (2006).
Levine, B. & Klionsky, D. *Dev. Cell* **6,** 463–477 (2004).
Lum, J. J., DeBarardinis, R. J. & Thompson, C. B. *Nature Rev. Mol. Cell Biol.* **6,** 439–448 (2005).
Ogier-Denis, E. & Codogno, P. *Biochim. Biophys. Acta* **1603,** 113–128 (2003).

REVIEWS

Multiple molecular mechanisms for multidrug resistance transporters

Christopher F. Higgins[1]†

The acquisition of multidrug resistance is a serious impediment to improved healthcare. Multidrug resistance is most frequently due to active transporters that pump a broad spectrum of chemically distinct, cytotoxic molecules out of cells, including antibiotics, antimalarials, herbicides and cancer chemotherapeutics in humans. The paradigm multidrug transporter, mammalian P-glycoprotein, was identified 30 years ago. Nonetheless, success in overcoming or circumventing multidrug resistance in a clinical setting has been modest. Recent structural and biochemical data for several multidrug transporters now provide mechanistic insights into how they work. Organisms have evolved several elegant solutions to ridding the cell of such cytotoxic compounds. Answers are emerging to questions such as how multispecificity for different drugs is achieved, why multidrug resistance arises so readily, and what chance there is of devising a clinical solution.

Since the discovery of effective antibiotics in the 1940s, a false sense of security has pervaded the public consciousness: the assumption that new drugs will increasingly conquer disease. This has proved over-optimistic, nowhere more so than the re-emergence of antibiotic-resistant infections such as tuberculosis. The problem is not restricted to antimicrobials—around 40% of human tumours develop resistance to chemotherapeutic drugs. With hindsight, it is not surprising that drug resistance is selected and spreads rapidly through cell populations. What is, perhaps, surprising is the phenomenon of multidrug resistance—the simultaneous acquisition of resistance to many chemically unrelated compounds to which the cell has never been exposed. Multidrug resistance is, in large part, the story of membrane transporters.

Cellular resistance to a single class of cytotoxic drugs can arise in many ways, including alteration of the target protein, decreased membrane permeability and drug metabolism. In contrast, the principal mechanism of multidrug resistance is the active transport of drugs out of the cell. Typically, each active transport protein is highly specific for its substrate, be it an amino acid, sugar or polypeptide. Unusually, however, multidrug transporters have broad specificity for a wide range of chemically unrelated molecules. Multidrug transporters, whether from *Escherichia coli* or an elephant, have similar (but not identical) multispecificity for many relatively lipophilic, planar molecules of molecular weight less than around 800 Da that are often, but not exclusively, weakly cationic (Fig. 1). These characteristics mirror those of many biologically active drugs and it is, therefore, not surprising that multidrug transporters in humans influence drug delivery and pharmacokinetics[1].

Active membrane transporters, whatever their substrate, fall into a relatively small number of protein superfamilies. Transporters within each superfamily are related with regards to amino acid sequence, structure and evolutionary origin. Intriguingly, multidrug transporters occur within several of these superfamilies[2] and must, therefore, have evolved several times, independently, in the context of very different protein backbones. These transporters present many intellectual and experimental challenges. How do they pump lipophilic drugs vectorially across lipid membranes? How can the very different architectures of different families of transporter each be adapted to multidrug transport? How is multispecificity achieved? What is the normal physiological function of these transporters?

Structures have now been obtained for multidrug transporters from four distinct transporter superfamilies (Fig. 2): (1) the ABC family (ATP-binding cassette: Sav1866 from *Staphylococcus* and mammalian P-glycoprotein); (2) the MFS family (major facilitator superfamily: EmrD from *E. coli*); (3) the RND family (resistance-nodulation-division: AcrB from *E. coli*); (4) the SMR family (small multidrug resistance: EmrE from *E. coli*). Together with structures of soluble, drug-binding transcription factors these data now provide satisfying, if incomplete, insights into the mechanisms and biology of multidrug resistance.

ABC transporters: Sav1866 and P-glycoprotein

ATP-binding cassette (ABC) transporters are present in all cells of all organisms and use the energy of ATP binding/hydrolysis to transport

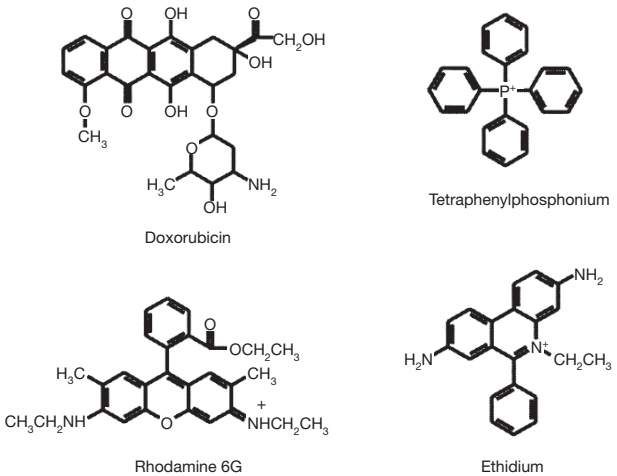


Figure 1 | Substrates of multidrug transporters. Multidrug transporters have subtly different multispecificities. However, their substrates share a number of features in common: planar, heterocyclic, lipophilic compounds of molecular mass less than 800 Da and, often, weakly cationic.

[1]MRC Clinical Sciences Centre, Imperial College, Hammersmith Hospital Campus, Du Cane Road, London W12 0NN, UK.
†Present Address: Vice-Chancellor's Office, Durham University, The University Offices, Old Elvet, Durham DH1 3HP, UK.

substrates across cell membranes[3]. Typically, they are specific for a given ligand that can be an inorganic ion, amino acid, sugar, polypeptide, or any one of a number of other classes of molecule. However, a few ABC transporters have evolved a broad specificity for hydrophobic molecules. Mammalian P-glycoprotein (ABCB1) is, arguably, the best characterized of all ABC transporters and, when overexpressed, confers resistance of cancer cells to a variety of chemotherapeutic drugs (for example, doxorubicin, Taxol, etoposide)[4,5]. Multidrug ABC transporters have also been implicated in antibiotic resistance, drug resistance in fungi and parasitic protozoa, and herbicide resistance in plants[6]. Intriguingly, the bacterial transporter LmrA, when expressed in mammalian cells, confers multidrug resistance indistinguishable from that of mammalian P-glycoprotein[7]. The minimal functional unit of all ABC transporters consists of four domains[8] (Fig. 2). Two cytoplasmic, nucleotide-binding domains (NBDs) bind and hydrolyse ATP and share a common protein fold distinct from that of other ATP-binding proteins. Two transmembrane domains (TMDs) each consist of multiple (generally six) membrane-spanning α-helices and form the pathway through which substrates cross the membrane. These four domains can be fused into multidomain polypeptides in a variety of ways. Bacterial multidrug transporters (for example, Sav1866) are most commonly homodimers of molecules comprising one NBD and one TMD, whereas mammalian P-glycoprotein has all four domains fused into a single polypeptide.

P-glycoprotein was the first multidrug transporter for which structural data were obtained, albeit at low-to-medium resolution[9–12]. These remain the only structural data for any mammalian multidrug transporter. Recently, a high-resolution structure of a homologous bacterial multidrug ABC transporter, Sav1866 from *Staphylococcus aureus*, was determined[13] (Fig. 3). The Sav1866 structure is consistent with the lower resolution structures[9–12] and biochemical cross-linking data[14,15] for P-glycoprotein. The two TMDs form a chamber in the membrane which, at least in the equivalent to the ATP-bound state (see below), is open extracellularly. This chamber is lined by hydrophobic and aromatic amino acids contributed by several transmembrane α-helices. The two NBDs form a head-to-tail 'sandwich' dimer in the intact protein, aligned such that each NBD contacts both TMDs. Two ATP-binding pockets are formed at the NBD dimer interface with amino acids from each monomer contributing to each ATP-binding pocket[16].

Structures for two substrate-specific ABC transporters from bacteria have also been determined: the vitamin B_{12} transporter BtuCD[17] and a metal-chelate transporter, HI1470-1 (ref. 18). Three putative structures for MsbA, a lipid A transporter, have been retracted[19]. As expected for homologous proteins, the structures of BtuCD and HI1470-1 are closely related to each other, and each transporter has a total of 20 transmembrane α-helices. However, although their overall architecture is similar to that of Sav1866/P-glycoprotein, there are significant differences in detail. The NBD dimer is similar in the BtuCD/HI1470-1 and Sav1866/P-glycoprotein pairs but the TMDs have unrelated folds. Furthermore, in BtuCD/HI1470-1 each NBD contacts only one of the two TMDs, whereas in Sav1866/P-glycoprotein each NBD contacts both TMDs. Biochemical cross-linking for P-glycoprotein confirms that each NBD interacts with two TMDs in this sub-group of proteins[20]. Thus, it seems that the two families of ABC transporters—the ABCB sub-family (Sav1866 and P-glycoprotein) and the BtuCD/HI1470-1 sub-family—have similar NBD dimers coupled to structurally and evolutionarily distinct pairs of TMDs. Consistent with this, ABC dimers are also known to couple ATP binding/hydrolysis to other very different classes of protein including DNA repair enzymes[21].

These structural data, together with extensive biochemical and genetic characterization, have led to the ATP-switch model for transport[21]. The driving force for drug transport is a switch between two principal conformations of the NBD dimer: ATP binding induces rigid body rotation of domains within each NBD with respect to each other and formation of a closed dimer with two molecules of ATP sandwiched at the dimer interface. ATP hydrolysis and inorganic phosphate (P_i)/ADP release return the dimer to its open configuration. The close proximity of the two NBDs in structures of intact ABC transporters suggests that the structural differences between the open and closed dimers are probably subtle rather than complete dimer dissociation. The kinetics of the switch can differ between transporters depending on the extent of cooperativity between the two nucleotide-binding pockets and signals from the transmembrane domains. ATP-binding by the NBDs and formation of the closed dimer induce substantial conformational changes in the TMDs[10,22] that mediate substrate translocation—a reduction in the drug-binding affinity[23–26] and reorientation of the binding site so that it is exposed to the extracellular face of the membrane[27] and drug can be released. ATP hydrolysis and P_i/ADP release restore the open dimer and the transporter to its starting configuration. The functional role of ATP binding and closed dimer formation is illustrated by studies of another ABC protein, CFTR, where, instead of mediating transport the ATP switch opens a chloride channel[28]. The nature of the conformational changes in the TMDs of ABC transporters is unknown. However, biochemical data, comparison of the ATP-bound and ATP-free forms of P-glycoprotein, and comparison of the BtuCD/HI1470-1 pair—which are thought to be in the ATP-bound (closed dimer) and ATP-free (open dimer) conformations, respectively—imply relatively small-scale tilting and rotation of several individual transmembrane α-helices with respect to each other to expose alternately the central chamber and drug-binding site(s) to the extracellular and cytoplasmic faces of the membrane[21].

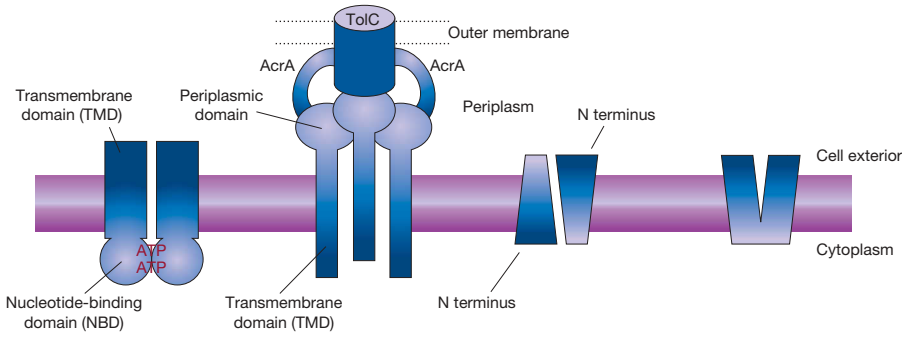


Figure 2 | Schematic diagram of domain organization of multidrug transporters. Examples of each of the four major families of transporters that include multidrug transporters and for which structural data are available.

As no structures of ABC transporters with bound substrate have been obtained, the nature of the substrate-binding site(s) can only be inferred. Drugs bind to a high-affinity site(s) on the protein from the inner leaflet of the lipid bilayer[29,30]. Mutations that alter drug-binding specificity of P-glycoprotein implicate several α-helices that line the central chamber[21,31]. It is reasonable, therefore, to suppose that the drug-binding site(s) is located in this chamber. Competitive and non-competitive drug-binding interactions[23,32], the observation that one drug can stimulate transport of another[33,34], and the demonstration that two drug molecules can bind per protein molecule[27,35], implicate multiple drug-binding sites. However, these data are also compatible with a single, large, flexible pocket that can bind more than one drug molecule simultaneously. On the basis of data for other multidrug-binding proteins this now seems to be the most plausible hypothesis (see below).

RND transporters: AcrB

Resistance-nodulation-division (RND) proteins are found in both prokaryotic and eukaryotic cells and have diverse substrate specificities and physiological roles. However, there are relatively few RND transporters and they are secondary transporters, energized not by ATP binding/hydrolysis but by proton movement down the transmembrane electrochemical gradient. Few RND proteins have been well characterized. The family includes NPC-1, which is defective in Niemann–Pick disease and modulates subcellular lipid/cholesterol distribution[36], and Dispatched in *Drosophila* and mammals which is required for the export of the cholesterol-modified signalling peptide Hedgehog[37]. Neither NPC-1 nor Dispatched, however, have formally been demonstrated to transport their putative substrates. By far the best characterized RND protein is AcrB from *E. coli* that can increase resistance to a variety of antibiotics by several orders of magnitude[38]. Recent structures of AcrB provide insights into the mechanism by which it, and presumably other RND transporters, works[39–42].

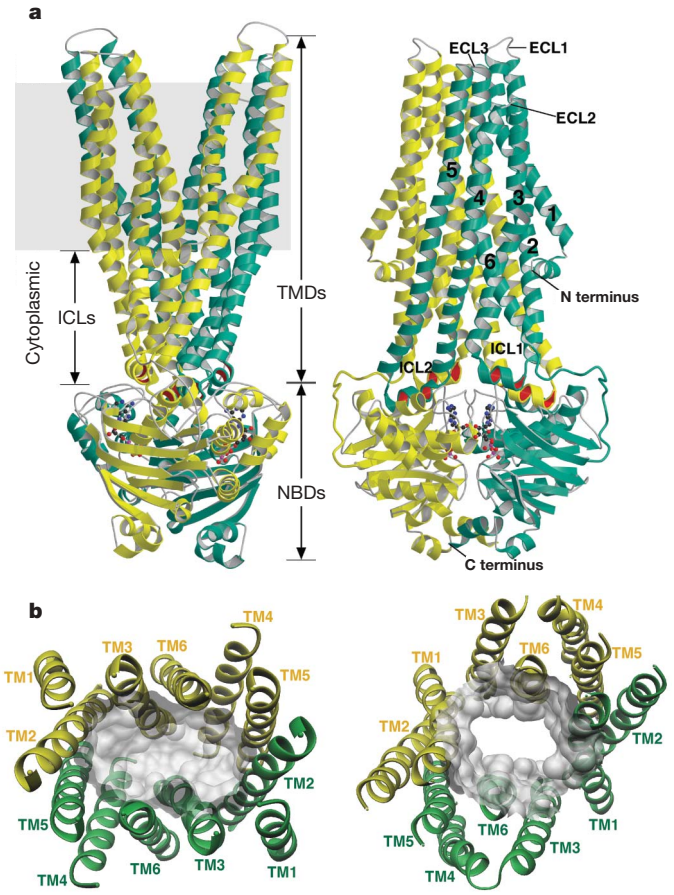


Figure 3 | Structure of ABC multidrug transporters. The backbone structure of Sav1866 is shown in ribbon representation. Sav1833 is a homodimer and the two monomers are coloured yellow and turquoise. **a**, View perpendicular to the cell membrane, in two orientations at right angles to each other. The TMDs span the lipid bilayer and consist of a total of 12 membrane-spanning α-helices. The NBDs are exposed at the cytoplasmic face of the membrane, linked to the TMDs by intracellular loops (ICLs). The six transmembrane α-helices of one subunit are numbered. The grey box indicates the probable position of the lipid membrane bilayer. **b**, View in the plane of the membrane showing the substrate translocation pathway, from the intracellular (left panel) and extracellular (right panel) faces of the membrane. The transmembrane (TM) helices are numbered and the cavity is shown as grey shading. Figure adapted from ref. 13.

Gram-negative bacteria including *E. coli* have two cell membranes—a cytoplasmic membrane and an outer membrane—separated by the periplasmic space. Many antibiotic targets are located in the periplasmic space (for example, the cell wall components targeted by β-lactam antibiotics). Thus, to confer resistance against a broad spectrum of antibiotics, active transporters must not only pump them out of the cytoplasm but also across the outer membrane. As the outer membrane is unable to maintain an electrochemical gradient or access ATP, energy input requires proteins located in the cytoplasmic membrane. AcrB is one such transporter (Fig. 2). AcrB consists of a transmembrane domain with 12 membrane-spanning α-helices and a large periplasmic domain. The functional transporter is a trimer with a total of 36 membrane-spanning α-helices. Two 'helper' proteins, AcrA and TolC, are required for AcrB to pump antibiotics out of the cell. AcrA is thought to have a role in membrane fusion but also has a more active, but poorly understood, role in the transport event itself[38,43]. TolC is a pore-like molecule comprising a 100 Å α-helical pore that spans the periplasm and a 40 Å β-barrel that spans the outer membrane[44]. AcrB translocates drugs into the TolC pore through which they cross both the periplasm and outer membrane. As TolC can couple to many different transporters, besides AcrB, it serves a generic role and has little or no function in determining the specificity or directionality of transport.

The structure of the AcrB trimer is shown in Fig. 4. The periplasmic domain adopts a subtly different conformation in each subunit of the trimer[40,42]. Each contains a potential substrate-recognition site and structures with bound drugs have been obtained[40]. However, only one site (the binding site; Fig. 4) is occupied by substrate at any given time. A second site (the extrusion site) is closed to the periplasm but open to the TolC docking domain, suggesting that it has just released substrate into the TolC pathway to exit the cell. In this site the drug pocket is smaller and the phenylalanine side chains are realigned so drug cannot enter or form stacking interactions—exactly as predicted for a low-affinity 'release' site. The third site (the access site) is closed to the TolC pathway but open to the periplasm, apparently ready to accept substrate. It does not take much imagination to envisage an 'alternating sites' model in which each of the three periplasmic domains adopts each of the three conformations in turn, passing substrate through the periplasmic domains to the TolC pathway and out of the cell. The movement of substrate through each periplasmic domain has been described as peristaltic[42]. Although analogous to the F_1F_0 ATPase, AcrB has no rotating subunit and the conformational changes must be induced directly by protons passing across the membrane down their electrochemical gradient. Three charged residues (Asp 407, Asp 408 and Lys 940) in the transmembrane domains that are conserved among all RND proteins and essential for function[39] probably mediate proton movement. Thus, like ABC transporters, energy transduction mediated by one domain (the transmembrane domain) is transduced by way of a conformational change to a second domain (the periplasmic domain) that mediates vectorial drug transport through changes in the affinity and orientation of a substrate-binding site. Without further data the molecular basis of energy transduction remains obscure.

As for other multidrug transporters, AcrB transports a plethora of hydrophobic compounds out of the cell. AcrB has been crystallized, separately, with two different bound substrates: minocycline and doxorubicin[40]. Despite their chemical differences both drugs bind in the same cavity in the periplasmic domain. This cavity is lined by hydrophobic and aromatic amino acids but also includes two polar residues, Gln 176 and Asn 274, that help to neutralize the charge of cationic drugs. The two drugs bind in different although overlapping places within this cavity and interact with different amino acid side chains. The finding that the substrate-binding site is in the periplasmic domain implies that AcrB 'picks up' drug from the outer (periplasmic) leaflet of the membrane to transport it out of the cell. This raises an apparent paradox: does AcrB also transport drugs from the cytoplasm? It is possible that another transporter facilitates drug 'flopping' from the inner to the outer leaflet from which AcrB pumps drug out of the cell. However, the three transmembrane domains of AcrB form a large (30 Å) hydrophobic cavity that appears to span much of the lipid bilayer and was originally suspected to form the drug translocation pathway. Indeed, an early structure showed three ligand molecules bound in this pathway[41]. Although subsequent data are inconsistent with these being sites from which transport occurs, it seems likely that AcrB also transports drugs directly from the cytoplasm, because a homologous protein, AcrD, has been shown to transport aminoglycosides from the cytoplasm[44].

SMR transporters: EmrE

Small multidrug resistance (SMR) proteins are a relatively small family of transporters, restricted to prokaryotic cells. They are also the smallest multidrug transporters, with only four transmembrane α-helices and no significant extramembrane domain, although as they function as dimers the minimal functional unit is a bundle of eight α-helices (Fig. 2). The paradigm SMR transporter, EmrE, is an electrogenic antiporter from *E. coli* that can confer resistance to a wide variety of hydrophobic cationic molecules, including antibiotics[45].

A structure obtained by cryo-electron microscopy to 7 Å resolution[46], together with genetic and biochemical data[47], shows that EmrE is a homodimer (Fig. 5). Two putative X-ray structures that were inconsistent with these data have recently been retracted[19]. Unusually, the two identical subunits appear to be oriented oppositely (antiparallel) in the membrane, although the folds of each monomer are subtly different. Although an antiparallel arrangement has been challenged by protein cross-linking studies[48], antiparallel homodimers and heterodimers are increasingly being recognized among membrane proteins and evolutionary and topological mapping now support an antiparallel arrangement for the subunits of EmrE[49].

EmrE is, in essence, a simple bundle of eight transmembrane α-helices that forms a pathway across the membrane. This pathway is lined primarily by hydrophobic and aromatic amino acids from several of the α-helices and is potentially accessible from both sides of the membrane (Fig. 5). The structure with bound TPP^+ (tetraphenylphosphonium) shows a relatively large, open pocket within the core of the hydrophobic transmembrane pathway and close to the middle of the membrane bilayer[46]. The involvement of multiple α-helices in forming the binding site/translocation pathway is consistent with extensive mutagenesis and other studies[47]. Glu 14, known to be essential for binding cationic drugs, is appropriately located. This residue has also been implicated in proton movement and, although the mechanism of proton coupling remains unknown, the involvement of Glu 14 in both substrate- and proton-binding suggests that the mechanisms of energy coupling in this family may differ from RND and MFS transporters[50]. The precise nature of the drug-binding site and the proton-induced conformational changes that presumably expose the site alternately to opposite faces of the membrane to achieve transport, remain unknown.

MFS transporters: EmrD

Major facilitator superfamily (MFS) transporters and ABC transporters comprise the two largest and most functionally diverse of the transporter superfamilies. However, MFS transporters are distinct from ABC transporters in both their primary sequence and structure (Fig. 2) and in the mechanism of energy coupling. As secondary transporters they are, like RND and SMR transporters, energized by the electrochemical proton gradient. Only in 2003 were the first X-ray structures for MFS transporters determined, for LacY[51] and GlpT[52] from *E. coli*. Subsequently, the structure of a single multidrug transporter from this superfamily, EmrD from *E. coli*, was determined to 3.5 Å resolution[53]. EmrD extrudes a range of cytotoxic molecules from the cell, although it is otherwise not well characterized. Fortunately, however, EmrD is homologous to two other MFS multidrug transporters that have been characterized biochemically in some detail: LmrP from *Lactococcus lactis* and MdfA from *E. coli*.

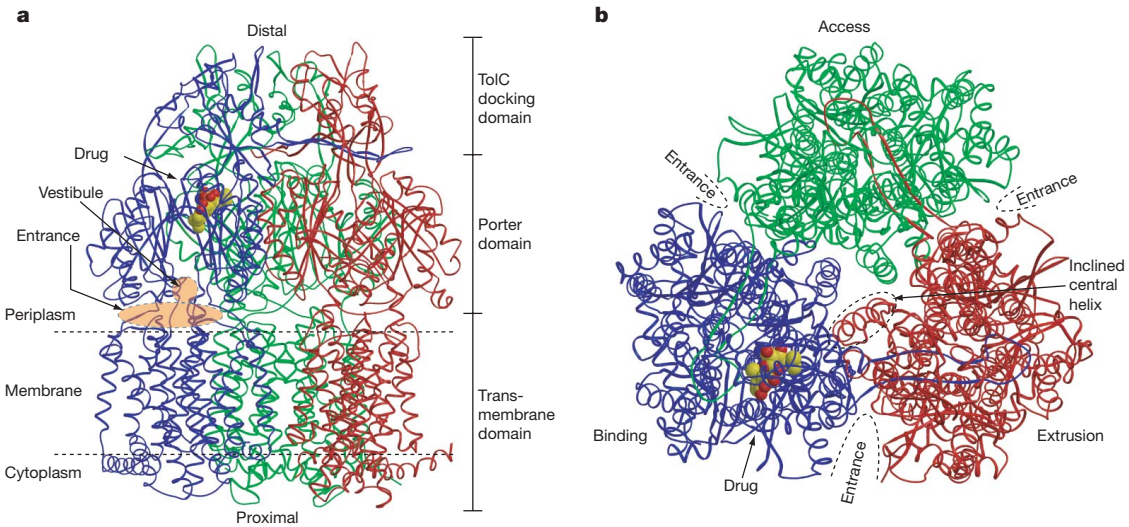


Figure 4 | Structure of RND multidrug transporters. The structure of AcrB in complex with minocycline is shown. AcrB is a trimer and the three monomers are coloured blue, red and green. The bound substrate minocycline is shown in the periplasmic domain as coloured balls and is present in one of the monomers only. **a**, View perpendicular to the membrane plane. The approximate position of the membrane bilayer is indicated. **b**, View parallel to the membrane plane, from the periplasmic side. The three periplasmic domains differ subtly in structure and the 'binding', 'access' and 'extrusion' subunits are indicated. Figure adapted from ref. 40.

EmrD is a compact protein with twelve membrane-spanning α-helices organized as two bundles of six that form a hydrophobic cavity within the plane of the bilayer (Fig. 5). The external α-helices (helices 3, 6, 9, 12) adopt a similar configuration to their equivalents in LacY and GlpT. The more internal α-helices deviate in their arrangement and form a larger internal cavity, presumably because of the different substrate specificities—LacY and GlpT are very substrate-specific whereas EmrD has broad multispecificity. This suggests that this internal cavity forms part of the drug transport pathway. No structure with bound drug has yet been obtained and there is no direct evidence that drugs bind in, or are transported through, this cavity. Nevertheless, mutational data for MdfA[54] and LmrP[55] imply that residues lining this central cavity are indeed involved in substrate recognition and translocation. These are, primarily, hydrophobic and aromatic amino acids but also include some polar residues. Notably, mutagenesis data suggest that many residues contribute to substrate binding, with different residues being more or less significant for different substrates. A good example is E26, which is important for the transport of cationic drugs but much less so for neutral ones[54]. Different drugs show complex competitive and non-competitive interactions with LmrP, leading to the suggestion that there may be multiple binding sites[56]. Indeed, MdfA appears to be able to bind chloramphenicol and TPP^+ simultaneously[57]. However, as for P-glycoprotein (see above), these data are also compatible with the simpler interpretation that there is a single large and flexible drug-binding site (see below).

Apart from its structure and the known requirement for the electrochemical gradient, little is known about the mechanisms of energy coupling or drug transport by EmrD. However, the related LacY protein is, arguably, the most intensively studied of all transporters[58] and a clear kinetic model has been established. For each molecule of lactose transported, a proton is transferred across the membrane via conserved and essential acidic residues. The proton and substrate (lactose) pathways seem to be distinct. Proton movement induces a conformational change that exposes the lactose-binding site to the external face of the membrane, reducing the affinity for lactose binding and facilitating its consequent release. The proton is then released and the transporter returns to its basal state with a high-affinity lactose-binding site exposed to the cytoplasm. It is likely that the principles established for LacY apply to the less-extensively studied drug transporters. However, there may be some adaptations to enable drugs of different charge to be accommodated. For example, in LmrP the two acidic residues (D142 and E327) involved in proton translocation are not essential but are individually replaceable, influencing the proton/substrate stoichiometry[55]. The relative importance of $\Delta\psi$ and ΔpH also seems to depend on the charge of the drug. As no structures of an MFS transporter at different stages of the transport cycle have yet been determined, the conformational changes that occur during transport can only be speculative.

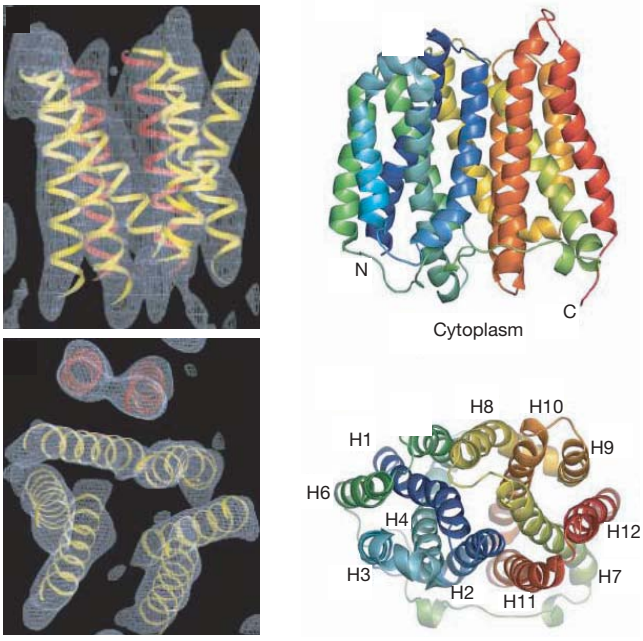


Figure 5 | Structures of SMR and MFS multidrug transporters. Left panels: structure of EmrE from *E. coli* determined by cryo-electron microscopy, viewed perpendicular to the membrane plane (top) and parallel to the membrane plane (bottom). The two antiparallel subunits each have four membrane-spanning α-helices. Taken, with permission, from ref. 46. Right panels: structure of EmrD from *E. coli*, viewed perpendicular (top) and parallel (bottom) to the plane of the membrane from the cytoplasmic side. The 12 membrane-spanning α-helices are numbered. Taken, with permission, from ref. 53.

Drug-binding transcription factors: QacR and BmrR

Structures of multidrug transporters with bound drug are limited. Nevertheless, important insights into the nature of multispecific drug-binding sites have been gained from studies of multidrug-binding transcription factors. In bacteria, expression of several multidrug transporter genes is induced by their cognate drug substrates. Soluble transcription factors bind drug and mediate this response, and the multispecificity of drug binding by these transcription factors is, unsurprisingly, similar to that of their partner transporters. These transcription factors have been much more amenable to structural study than the membrane-bound transporters. Elegant studies of two of them have been particularly informative: BmrR from *Bacillus subtilis*[59,60] and QacR from *Staphylococcus aureus*[61,62]. Structures with and without bound drug show that, although the folds of their multidrug-binding domains differ, the two proteins bind drugs in a similar manner.

BmrR and QacR each have a relatively large drug-binding pocket that can accommodate the entire spectrum of drug ligands—there is no need for multiple pockets to explain multispecificity[63]. The architecture of the drug-binding pockets allows different ligands to adopt different orientations within the pocket and interact with different sets of amino acids. For example, in QacR two chemically diverse ligands (rhodamine and ethidium) occupy distinct, almost non-overlapping sites within the binding pocket and interact with different amino acid side chains. A structure has also been obtained in which two different drugs (ethidium and proflavin) are bound simultaneously[62] (Fig. 6). Regions of the pocket not occupied by ligand are occupied by water molecules, as initially described for the polyspecific oligopeptide-binding protein of *Salmonella typhimurium*[64]. The pocket wall is flexible and can change conformation upon ligand binding, increasing promiscuity. Nevertheless, flexibility is limited, not unexpectedly for a folded protein, explaining why addition of a specific side chain to some drugs can reduce binding affinity.

The drug-binding pockets shield bound drug from the aqueous phase and drug binding is stabilized by van der Waals interactions with the surrounding hydrophobic and aromatic amino acid side chains. Sequestration of drug in a hydrophobic pocket provides sufficient energy to stabilize binding, as it negates the disruption of hydrogen bonds between water molecules otherwise caused by drug molecules in aqueous solution. Binding affinity for cationic drugs is augmented by electrostatic attraction between the positively charged ligand and negative charges on the protein. The hydrophobic environment of the drug-binding pocket makes electrostatic attraction an especially powerful stabilizing factor as water dipoles are avoided. Notably, this electrostatic interaction does not require perfect alignment of the positive and negative charges. For example, for BmrR the closer the positive charge of the bound drug to the single glutamate (Glu 134) in the binding pocket the higher the drug-binding affinity[59]. In QacR, the four glutamates exposed in the drug-binding pocket stabilize the binding of different drugs, depending on how the specific drug is aligned in the pocket[61]. In the unliganded proteins

the charged residues are shielded from the aqueous phase, for example in BmrR by a flexible α-helical arm that is displaced when drug binds[59].

Although not linked to transport, the structures of three other proteins that bind multiple hydrophobic ligands illustrate a similar mechanism of multispecific binding[63]. The drug-metabolizing enzyme P450 not only has a binding site with similar characteristics to those of BmrR and QacR, but a structure has been obtained with two identical drug molecules in the same pocket, explaining homotopic cooperativity[65]. The PXR transcriptional regulator[66] has a similar binding site, as does the mammalian odorant-binding protein, which, intriguingly, is constructed from β-sheets rather than α-helices[67].

Insights from comparison of multidrug transporters

It has proved unusually difficult to obtain structures of drug transporters, in part owing to the general problems of purifying and crystallizing membrane proteins. Multispecific transporters may be particularly problematical as they are required to be flexible in order to translocate relatively large molecules across the bilayer—flexibility can lead to anomalous crystal contacts when the protein is removed from the lipid environment. This is illustrated by two of the first multidrug transporter X-ray structures obtained (MsbA[68] and EmrE[69] from *E. coli*), which, even disregarding software difficulties[19], appear to be anomalously folded proteins. Thus, once a structure has been determined it should not necessarily be assumed to reflect the physiological fold. Structure determination in the absence of demonstrable biochemical activity should be strenuously avoided. More importantly, it is critical that multiple structures are obtained, compared and tested against independent biochemical and genetic data to give a level of reassurance about biological relevance before more detailed mechanistic interpretation.

As we have seen, multidrug transporters have evolved several times, apparently independently, on very different protein backbones. These transporters perform similar functions yet achieve this by very different means. Nevertheless, several common principles emerge.

First, multidrug transporters are conventional enzymes that can be exemplified by simple kinetic schemes, as first proposed by Mitchell nearly 50 years ago[70]. Although this now seems self-evident, the unusual properties of multidrug transporters have led to alternative speculations as to how drugs cross the membrane, including formation of discontinuities in the lipid bilayer and 'slippery' protein–lipid interfaces. We do not yet have a complete set of structures for any transporter at different stages in the transport cycle, but the current body of structural and biochemical data, *in toto*, shows that substrates are bound by a defined high-affinity site exposed to one face of the membrane. Conformational changes induced by ATP binding/hydrolysis or proton movement down the electrochemical gradient convert this site to a low-affinity 'release' site exposed to the alternative face of the membrane. These conformational changes can be within a protein domain (as for MFS and SMR transporters) or transmitted between domains (as for ABC and RND transporters).

Second, substrates cross the bilayer through a pathway formed within the core of the transporter, largely shielded from the surrounding lipid phase. For some multidrug resistance transporters the evidence comes directly from structures with bound drug, although for others it is based on indirect mutagenesis studies. This finding is, perhaps, not surprising, as most multidrug resistance transporters are closely related to transporters that transport hydrophilic substrates that must clearly be shielded from the lipid phase. The pathway is not strictly a channel, as it is not open to both faces of the membrane simultaneously but instead alternately during the transport cycle. In all transporters the pathway is constructed from

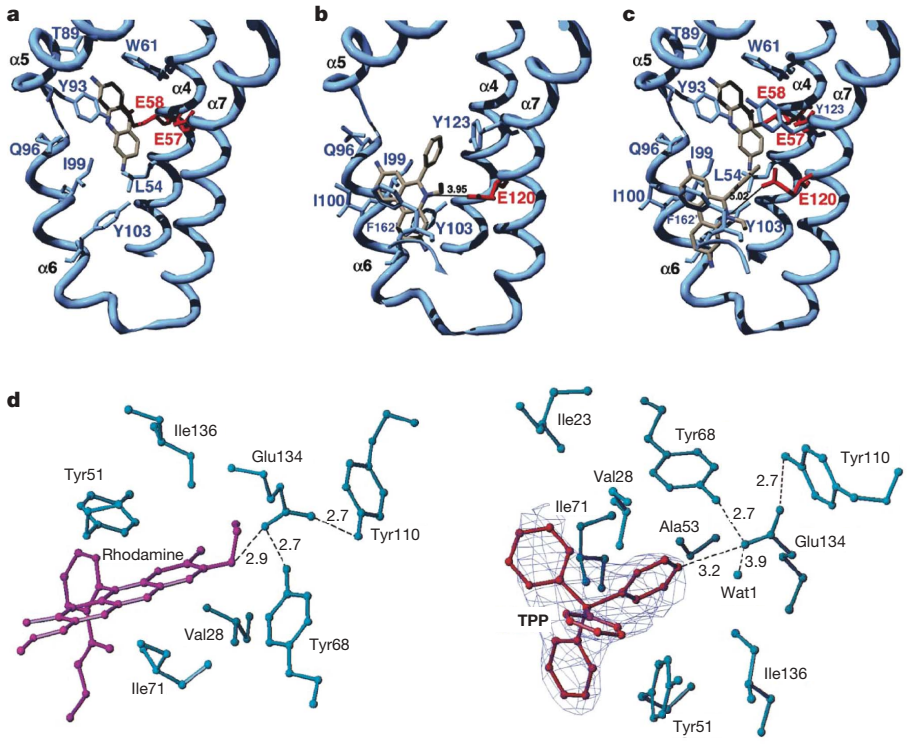


Figure 6 | Drug binding by soluble bacterial transcription factors. **a**–**c**, Binding of three different drugs by QacR. **a**, QacR with bound proflavin; **b**, QacR with bound ethidium; **c**, QacR with both proflavin and ethidium bound simultaneously in a tertiary complex. Only the key drug-binding residues are shown. Acidic residues involved in neutralization of cationic drugs are shown in red. Taken, with permission, from ref. 62. **d**, Drug binding by BmrR. The key residues involved in drug binding are shown in ball-and-stick configuration. Left: structure-based model with bound rhodamine (purple balls and sticks). The key neutralizing acidic residue, Glu 134, is shown. Right: structure with bound TPP^+ (in red) together with a bound water molecule (Wat 1). Values shown are in angstrom. Dashed lines indicate H bonds. Taken, with permission, from ref. 59.

multiple α-helices. However, because the TolC pore through which drugs cross the bacterial outer membrane is formed by β-sheets—as is the drug-binding site of the olfactory receptor—the involvement of α-helices may say more about the constraints of building a membrane transporter than anything profound about the specific requirements for a drug transport pathway per se.

Third, multidrug transporters extract their substrates from the inner leaflet of the bilayer, analogous to phospholipid flippases/floppases, which generate lipid asymmetry in the membrane[29]. AcrB is the exception that proves this rule, accessing drugs from the outer leaflet of the bilayer in order to remove them from the periplasm. Given this, it is perhaps not unexpected that multidrug transporters are related to lipid flippases/floppases. For example, the most closely related transporter to P-glycoprotein is the phosphatidylcholine floppase ABCB4 (ref. 71). This does not mean that multidrug transporters are themselves lipid flippases/floppases—just that they are mechanistically similar. The acquisition of substrate from a specific leaflet of the membrane has a number of potential advantages. (1) Hydrophobic substrates partition into the bilayer and so are at increased concentration compared with the aqueous phase; (2) drugs in the membrane diffuse in two- rather than three-dimensions, facilitating interactions with the transporter; (3) extraction of substrate from the membrane ensures that broad-specificity transporters do not expel normal cellular constituents that remain in the cytoplasm; and (4) capturing drugs from the inner leaflet, before they enter the cytoplasm, is the most effective means of ensuring that they do not interact with their cytoplasmic target[72], although for drugs for which flip-flopping between leaflets is rate-limiting, extraction from the periplasmic leaflet is most efficient.

The fourth common principle is that structures of soluble drug-binding proteins demonstrate that multispecificity is achieved by a single, large, flexible hydrophobic pocket in which drug is essentially shielded from both lipid and aqueous phase. These pockets can, in some cases, accommodate two identical or different drug molecules at once. The pockets are lined primarily by hydrophobic/aromatic amino acids that bind drugs via van der Waals and stacking interactions. Polar residues can negate nearby positive charges of weakly cationic drugs. Critically, these multispecific drug-binding sites are very different from the usual 'lock-and-key' mechanisms of enzymes with hydrophilic substrates that rely on a perfect spatial alignment between ligand and side chains in the binding site to overcome the energetic disadvantage of disrupting hydrogen bonds between ligand and water. In contrast, simply the removal of hydrophobic drugs from the aqueous environment is energetically favourable and the multiple, weak interactions between a hydrophobic ligand and hydrophobic amino acids are sufficient to generate high affinity.

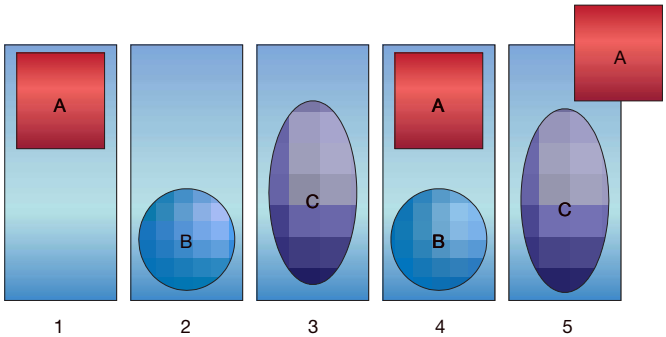


Figure 7 | One or multiple drug-binding sites? A single, large, drug-binding site can accommodate drug A (panel 1), drug B (panel 2) or drug C (panel 3). It can also bind drugs A and B simultaneously (panel 4), equivalent to two pharmacologically distinct sites, but is unable to bind drugs A and C simultaneously (panel 5), equivalent to a pharmacologically single site. MRP1 can bind drug and glutathione separately or as a conjugate[76,77], showing that the question of whether there are one or two sites is, to some extent, semantic.

The limited available structures of multidrug transporters with bound substrate (AcrB and EmrE) reveal drug-binding pockets that conform to this model. All other multidrug transporters have a similar hydrophobic pocket that appears to be the drug-binding site based on mutagenesis data. Although pharmacological and kinetic data showing both competitive and non-competitive drug interactions are consistent with multiple, interacting drug-binding sites, a single, large, flexible pocket can equally explain these data and now seems the most plausible model for all multidrug transporters (Fig. 7).

The biology of multidrug transporters

The ability of a cell to protect itself against environmental toxins is an essential biological function. Many organisms produce toxins to repel ecological competitors, and plants, which cannot run away from predators, rely on toxic secondary metabolites to make themselves unappetizing. Many antibiotics are derivatives of natural bacterial or fungal products, and many anticancer dugs (for example, Taxol, vinca alkaloids) are natural plant products. These molecules intercalate into lipid bilayers and are often deleterious to membrane function, altering fluidity, curvature or the activity of membrane proteins. To survive these natural chemical onslaughts, most organisms have evolved multidrug transporters to prevent cytotoxic molecules entering cells and to clear membranes of unwanted agents. This is self-evidently essential for *Streptomyces* strains that must protect antibiotic-producing cells from the very molecules they themselves synthesize. In mammals, the normal cellular function of P-glycoprotein is also protective, illustrated by the *mdr* (also known as *Abcb1b*) (P-glycoprotein) knockout mouse, which has no overt phenotype except altered drug pharmacokinetics and sensitivity to neurotoxins that are otherwise prevented from crossing the blood–brain barrier[2].

On the other hand, a plethora of proteins have been designated as multidrug transporters on the basis of sequence homologies, often with little or no biological evidence. In many cases this is likely to be a misnomer. Indeed, several proteins that transport and/or confer resistance to drugs when overexpressed have rather specific physiological roles when expressed at physiological levels. For example, MRP2 (ABCC2) can confer drug resistance when overexpressed yet is actually a leukotriene C_4 transporter[73]. The Blt 'multidrug transporter' of *Bacillus subtilis* confers drug resistance when overexpressed, yet its expression is normally co-induced with enzymes involved in spermidine/spermine metabolism and it is actually a spermidine transporter[74]. Overexpression of MdfA confers resistance of *E. coli* to many antibiotics, yet deletion of the *mdfA* gene barely alters cellular resistance. Instead, the real physiological role of MdfA turns out to be as a $Na^+(K^+)/H^+$ antiporter, which enables cells to maintain a constant intracellular pH under alkaline conditions[75].

It is now clear why drug resistance in nature most frequently involves multispecific membrane transporters. Resistance to a single drug can be achieved by mutation of its target. This is difficult to achieve without adversely altering the function of the target itself, and could not confer resistance to drugs with different targets. Similarly, enzymes designed to degrade a specific drug cannot metabolize unrelated molecules—57 cytochrome P450 genes are necessary for mammals to metabolize multiple cytotoxic drugs. Given the multiplicity of natural cytotoxic entities with different targets, the simplest means to achieve resistance is to take advantage of a common property of these molecules: the need to cross the cell membrane. Most natural toxins, like synthetic drugs, are small, planar, lipophilic molecules precisely because they have to cross the lipid bilayer to exert their toxic effects (with the rare exception of those that gain entry to the cell by 'piggy-backing' on a specific transporter or have an extracellular target). It turns out, as we have seen, that the evolution of a broad-specificity transporter for chemical entities with the characteristics required to cross the cell membrane is rather straightforward. This is the strategy adopted across the natural kingdom, and it is these natural resistance mechanisms that are frequently brought

into play when we try to interfere with nature by developing and using cytotoxic drugs.

Clinical implications

Over the past 20 years two general approaches have been adopted, relatively unsuccessfully, to overcoming drug resistance in the clinic. As discussed above, it is no coincidence that most clinically useful drugs are substrates for multidrug transporters. Given what we now know, what are the prospects of circumventing multidrug resistance in the clinic? Depending on perspective, views may either be 'glass-half-full' or 'glass-half-empty'.

First, there have been considerable efforts to modify drugs and antibiotics chemically so that they are no longer substrates for multidrug transporters. In the absence of structural data on the nature of multidrug-binding sites this has been undertaken 'blind' and has proved unsatisfying. Unlike conventional enzyme-substrate-binding sites, small changes in a drug rarely result in a substantial reduction in its affinity for transport. We now understand that this is because drug-binding sites are large and flexible and because the precise alignment of the ligand with respect to amino acids on the binding pocket is not required for high-affinity binding. Any modification to a drug that substantially reduces its affinity for a transporter also tends to reduce its 'druggability'—its ability to cross the cell membrane and bind to its target. It has, perhaps naively, been assumed that understanding the structure of multidrug-binding sites on transporters would enable drugs to be modified, rationally, to circumvent resistance. Instead, the current picture suggests that the very nature of drug-binding sites on transporters makes this a difficult proposition—certainly far more difficult than modifying hydrophilic ligands of cytosolic enzymes.

The second approach to overcoming multidrug resistance—the development of specific inhibitors of transporters—has also proved unsatisfactory[5]. There is frequently more than one multidrug transporter that can confer resistance to a specific drug or antibiotic, and so more than one inhibitor may be required. By trial and error, rather than rational design, several high-affinity and relatively specific inhibitors of human P-glycoprotein have been developed. These work well in the laboratory setting. However, they have proved difficult to assess in the clinic because of side-effects caused by inhibiting P-glycoprotein's normal physiological function in healthy tissues, altering the pharmacokinetics of the co-administered cytotoxic drug and enabling it to cross the blood–brain barrier.

Perhaps most worrying is the relative ease with which multidrug transporters can arise to confound our efforts at therapy. Most bacterial and mammalian cells have multidrug transporters to protect themselves against natural cytotoxic and membrane-disruptive compounds. It seems that selection for overexpression, or heterologous expression in different cell or tissue types, is relatively straightforward in response to a therapeutic or antimicrobial drug. Because of the nature of drug-binding sites, new multidrug transporters can evolve relatively easily from substrate-specific transporters by mutation. Indeed, some substrate-specific transporters can handle a variety of drugs without mutation, conferring multidrug resistance simply upon overexpression. Finally, cytotoxic drugs themselves upregulate multidrug transporters both in bacteria and in mammalian cells, often as part of a more general stress response.

Concluding remarks

Recent rapid advances in our understanding of multidrug transporters have not yet provided solutions to pressing clinical problems. These advances have, however, shown us why apparently straightforward approaches to overcoming multidrug resistance have been less effective than might otherwise have been expected. Increased understanding will inevitably enhance the chances of clever and effective solutions. In the meantime, profligate use of antibiotics has, unquestionably, led to the spread of antibiotic resistance. Similarly, chemotherapeutic drugs induce expression of drug resistance pathways. Should we perhaps stop assuming we can beat multidrug resistance and instead implement strategies to avoid it?

1. Schinkel, A. H., Wagenaar, E., Mol, C. A. & van Deemter, L. P-glycoprotein in the blood-brain barrier of mice influences the brain penetration and pharmacological activity of many drugs. *J. Clin. Invest.* **97,** 2517–2524 (1996).
2. Saier, M. H. & Paulsen, I. T. Phylogeny of multidrug transporters. *Semin. Cell Dev. Biol.* **12,** 205–213 (2001).
3. Higgins, C. F. ABC transporters: from microorganisms to man. *Annu. Rev. Cell Biol.* **8,** 67–113 (1992).
4. Juliano, R. L. & Ling, V. A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. *Biochim. Biophys. Acta* **455,** 152–162 (1976).
5. Gottesman, M. M. Fojo, T. & Bates, S. E. Multidrug resistance in cancer: role of ATP-dependent transporters. *Nature Rev. Cancer* **2,** 48–58 (2001).
6. Holland, I. B., Cole, S. P. C., Kuchler, K. & Higgins, C. F. (eds) *ABC Proteins: from Bacteria to Man* (Academic, London, 2003).
7. van Veen, H. W. *et al.* A bacterial antibiotic resistance gene that complements the human multidrug resistance P-glycoprotein gene. *Nature* **391,** 291–295 (1998).
8. Higgins, C. F. *et al.* A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. *Nature* **323,** 448–450 (1986).
9. Rosenberg, M. F., Callaghan, R., Ford, R. C. & Higgins, C. F. Structure of the multidrug-resistance P-glycoprotein to 2.5 nm resolution determined by electron microscopy. *J. Biol. Chem.* **272,** 10685–10694 (1997).
10. Rosenberg, M. F. *et al.* Repacking of the transmembrane domains of P-glycoprotein, during the transport ATPase cycle. *EMBO J.* **20,** 5615–5625 (2001).
11. Rosenberg, M. F., Kamis, A. B., Callaghan, R., Higgins, C. F. & Ford, R. C. Three-dimensional structures of the mammalian multidrug resistance P-glycoprotein demonstrate major conformational changes in the transmembrane domains upon nucleotide binding. *J. Biol. Chem.* **278,** 8294–8299 (2003).
12. Rosenberg, M. F., Callaghan, R., Modok, S., Higgins, C. F. & Ford, R. C. Three-dimensional structure of P-glycoprotein: the transmembrane regions adopt an asymmetric configuration in the nucleotide-bound state. *J. Biol. Chem.* **280,** 2857–2862 (2005).
13. Dawson, R. J. P. & Locher, K. P. Structure of a bacterial multidrug ABC transporter. *Nature* **443,** 180–185 (2006).
14. Loo, T. W. & Clarke, D. M. The packing of the transmembrane segments of human multidrug resistance P-glycoprotein is revealed by disulfide cross-linking analysis. *J. Biol. Chem.* **275,** 5253–5256 (2000).
15. Stenham, D. R. *et al.* An atomic detail model for the human ATP-binding cassette transporter, P-glycoprotein, derived from disulphide cross-linking and homology modelling. *FASEB J.* **17,** 2287–2289 (2003).
16. Smith, P. C. *et al.* ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol. Cell* **10,** 139–149 (2002).
17. Locher, K. P., Lee, A. T. & Rees, D. C. The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* **296,** 1091–1098 (2002).
18. Pinkett, H. W., Lee, A. T., Lum, P., Locher, K. P. & Rees, D. C. An inward facing conformation of a putative metal-chelate-type ABC transporter. *Science* **315,** 373–377 (2007); published online 7 December 2006.
19. Chang, G. *et al.* Retraction. *Science* **314,** 1875 (2006).
20. Zolnerciks, J. K., Wooding, C. & Linton, K. J. In vivo evidence for a Sav1866-like architecture for the human multidrug transporter P-glycoprotein. *J. Biol. Chem.* (submitted).
21. Higgins, C. F. & Linton, K. J. The ATP switch model for ABC transporters. *Nature Struct. Mol. Biol.* **11,** 918–926 (2004).
22. Dong, J., Yang, G. & Mchaourab, H. S. Structural basis for energy transduction in the transport cycle of MsbA. *Science* **308,** 1023–1028 (2005).
23. Dey, S., Ramachandra, M., Pastan, I., Gottesman, M. M. & Ambudkar, S. V. Evidence for two non-identical drug-interaction sites in the human P-glycoprotein. *Proc. Natl Acad. Sci. USA* **94,** 10594–10599 (1997).
24. Martin, C. *et al.* Drug binding sites on P-glycoprotein are altered by ATP binding prior to nucleotide hydrolysis. *Biochemistry* **39,** 11901–11906 (2000).
25. Martin, C., Higgins, C. F. & Callaghan, R. The vinblastine binding site adopts high- and low-affinity conformations during a transport cycle of P-glycoprotein. *Biochemistry* **40,** 15733–15742 (2001).
26. Payen, L. F., Gao, M., Westlake, C. J., Cole, S. P. C. & Deeley, R. G. Role of carboxylate residues adjacent to the conserved core walker B motifs in the catalytic cycle of multidrug resistance protein 1 (ABCC1). *J. Biol. Chem.* **278,** 38537–38547 (2003).
27. van Veen, H. W., Margolles, A., Muller, M., Higgins, C. F. & Konings, W. N. The homodimeric ATP binding cassette transporter LmrA mediates multidrug resistance by a two-site (two-cylinder engine) mechanism. *EMBO J.* **19,** 2503–2514 (2000).
28. Vergani, P., Lockless, S. W., Nairn, A. C. & Gadsby, D. C. CFTR channel opening by ATP-driven tight dimerization of its nucleotide binding domains. *Nature* **433,** 876–880 (2005).
29. Higgins, C. F. & Gottesman, M. M. Is the multidrug transporter a flippase? *Trends Biochem. Sci.* **17,** 18–21 (1992).
30. Bolhuis, H. *et al.* Multidrug resistance in *Lactococcus lactis*: evidence for ATP-dependent drug extrusion from the inner leaflet of the cytoplasmic membrane. *EMBO J.* **15,** 4239–4245 (1996).

31. Loo, T. W. & Clarke, D. M. Recent progress in understanding the mechanism of P-glycoprotein-mediated drug efflux. *J. Membr. Biol.* **206,** 173–185 (2005).
32. Martin, C. *et al.* Communication between multiple drug binding sites on P-glycoprotein. *Mol. Pharmacol.* **58,** 624–632 (2000).
33. Shapiro, A. B. & Ling, V. Positively cooperative sites for drug transport by P-glycoprotein with distinct drug specificities. *Eur. J. Biochem.* **250,** 130–137 (1997).
34. Zelcer, N. *et al.* Evidence for two interacting ligand binding sites in human multidrug resistance protein 2 (ABCC2). *J. Biol. Chem.* **278,** 23538–23544 (2003).
35. Lugo, M. R. & Sharom, F. J. Interaction of LDS-751 and rhodamine 123 with P-glycoprotein: evidence for simultaneous binding of both drugs. *Biochemistry* **44,** 14020–14029 (2005).
36. Ko, D. C., Gordon, M. D., Jin, J. Y. & Scott, M. P. Dynamic movements of organelles containing Niemann-Pick C1 protein: NPC1 involvement in late endocytic events. *Mol. Biol. Cell* **12,** 601–614 (2001).
37. Ma, Y. *et al.* Hedgehog-medicated patterning of the mammalian embryo requires transporter-like function of dispatched. *Cell* **111,** 63–75 (2002).
38. Zgurskaya, H. I. & Nikaido, H. Bypassing the periplasm: reconstitution of the AcrAB multidrug efflux pump of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **96,** 7190–7195 (1999).
39. Murakami, S., Nakashima, R., Yamashita, E. & Yamaguchi, A. Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* **419,** 587–593 (2002).
40. Murakami, S., Nakashima, R., Yamashita, E., Matsumoto, T. & Yamaguchi, A. Crystal structure of a multidrug transporter reveals a functionally rotating mechanism. *Nature* **443,** 173–179 (2006).
41. Yu, E. W., McDermott, G., Zgurskaya, H. I., Nikaido, H. & Koshland, D. E. Structural basis of multiple drug-binding capacity of the AcrB multidrug efflux pump. *Science* **300,** 976–980 (2003).
42. Seeger, M. A. *et al.* Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. *Science* **313,** 1295–1298 (2006).
43. Koronakis, V., Sharff, A., Koronakis, E., Luisi, B. & Hughes, C. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* **405,** 914–919 (2000).
44. Aires, J. R. & Nikaido, H. Aminoglycosides are captured from both periplasm and cytoplasm by the AcrD multidrug efflux transporter of *Escherichia coli*. *J. Bacteriol.* **187,** 1923–1929 (2005).
45. Schuldiner, S., Lebendiker, M. & Yerushalmi, H. EmrE, the smallest ion-coupled transporter, provides a unique paradigm for structure-function studies. *J. Exp. Biol.* **200,** 335–341 (1997).
46. Ubarretxena-Belandia, I., Baldwin, J. M., Schuldiner, S. & Tate, C. G. Three-dimensional structure of the bacterial multidrug transporter EmrE shows it is an asymmetric homodimer. *EMBO J.* **22,** 6175–6181 (2003).
47. Sharoni, M., Steiner-Mordoch, S. & Schuldiner, S. Exploring the binding domain of EmrE, the smallest multidrug transporter. *J. Biol. Chem.* **280,** 32849–32855 (2005).
48. Soskine, M., Mark, S., Tayer, N., Mizrachi, R. & Schuldiner, S. On parallel and antiparallel topology of a homodimeric multidrug transporter. *J. Biol. Chem.* **281,** 36205–36212 (2006).
49. Rapp, M., Granseth, E., Seppala, S. & von Heijne, G. Identification and evolution of dual-topology proteins. *Nature Struct. Mol. Biol.* **13,** 112–116 (2006).
50. Yerushalmi, H. & Schuldiner, S. A model for coupling of H^+ and substrate fluxes based on 'time-sharing' of a common binding site. *Biochemistry* **39,** 14711–14719 (2000).
51. Abramson, J. *et al.* Structure and mechanisms of the lactose permease of *Escherichia coli*. *Science* **301,** 610–615 (2003).
52. Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang, D. N. Structure and mechanisms of the glycerol-3-phosphate transporter from *Escherichia coli*. *Science* **301,** 616–620 (2003).
53. Yin, Y., He, X., Szewczyk, P., Nguyen, T. & Chang, G. Structure of the multidrug transporter EmrD from *Escherichia coli*. *Science* **312,** 741–744 (2006).
54. Adler, J. & Bibi, J. Promiscuity in the geometry of electrostatic interactions between the *Escherichia coli* multidrug resistance transporter MdfA and cationic substrates. *J. Biol. Chem.* **280,** 2721–2729 (2005).
55. Mazurkiewicz, P., Driessen, A. J. & Konings, W. N. Energetics of wild type and mutant multidrug resistance secondary transporter LmrP of *Lactococcus lactis*. *Biochim. Biophys. Acta* **1658,** 252–261 (2004).
56. Putman, M., Koole, L. A., van Veen, H. W. & Konings, W. N. The secondary multidrug transporter LmrP contains multiple drug interaction sites. *Biochemistry* **38,** 13900–13905 (1999).
57. Lewinson, O., Padan, E. & Bibi, E. Alkalitolerance: a biological function for a multidrug transporter in pH homeostasis. *Proc. Natl Acad. Sci. USA* **101,** 14073–14078 (2004).
58. Abramson, J., Iwata, S. & Kaback, H. R. Lactose permease as a paradigm for membrane transport proteins (Review). *Mol. Membr. Biol.* **21,** 227–236 (2004).
59. Zheleznova, E. E., Markham, P. N., Neyfakh, A. A. & Brennan, R. G. Structural basis of multidrug recognition by BmrR, a transcription activator of a multidrug transporter. *Cell* **96,** 353–362 (1999).
60. Vázquez-Laslop, N., Markham, P. N. & Neyfakh, A. A. Mechanisms of ligand recognition by BmrR, the multidrug-responding transcriptional regulator: mutational analysis of the ligand binding site. *Biochemistry* **38,** 16925–16931 (1999).
61. Schumacher, M. A. *et al.* Structural mechanisms of QacR induction and multidrug recognition. *Science* **294,** 2158–2163 (2001).
62. Schumacher, M. A., Miller, M. C. & Brennan, R. G. Structural mechanisms of the simultaneous binding of two drugs to a multidrug-binding protein. *EMBO J.* **23,** 2923–2930 (2004).
63. Neyfakh, A. A. Mystery of multidrug transporters: the answer can be simple. *Mol. Microbiol.* **44,** 1123–1130 (2002).
64. Tame, J. R. H. *et al.* The structural basis of sequence-independent peptide binding by OppA protein. *Science* **264,** 1578–1581 (1994).
65. Cupp-Vickery, J., Anderson, R. & Hatziris, Z. Crystal structures of ligand complexes of P450eryF exhibiting homotropic competition. *Proc. Natl Acad. Sci. USA* **97,** 3050–3055 (2000).
66. Watkins, R. E. *et al.* The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity. *Science* **292,** 2329–2333 (2001).
67. Ramoni, R. *et al.* The insect attractant 1-octen-3-ol is the natural ligand of bovine odorant-binding protein. *J. Biol. Chem.* **276,** 7150–7155 (2001).
68. Chang, G. & Roth, C. B. Structure of MsbA from *E. coli*: a homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science* **293,** 1793–1800 (2001).
69. Ma, C. & Chang, G. Structure of the multidrug resistance efflux transporter EmrE from *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **101,** 2852–2857 (2004).
70. Mitchell, P. A general theory for membrane transport from studies of bacteria. *Nature* **180,** 134–136 (1957).
71. Smit, J. J. M. *et al.* Homozygous disruption of the murine *mdr2* P-glycoprotein gene leads to a complete absence of phospholipid from bile and to liver disease. *Cell* **75,** 451–462 (1993).
72. Stein, W. D., Cardarelli, C., Pastan, I. & Gottesman, M. M. Kinetic evidence suggesting that the multidrug transporter differentially handles influx and efflux of its substrates. *Mol. Pharmacol.* **45,** 763–772 (1994).
73. Buchler, M. *et al.* cDNA cloning of the hepatocyte canalicular isoform of the multidrug resistance protein, cMRP, reveals a novel conjugate export pump deficient in the hyperbilirubinemic mutant rats. *J. Biol. Chem.* **271,** 15091–15098 (1996).
74. Ahmed, M. *et al.* Two highly similar multidrug transporters of *Bacillus subtilis* whose expression is differentially regulated. *J. Bacteriol.* **177,** 3904–3910 (1995).
75. Lewinson, O. & Bibi, E. Evidence for simultaneous binding of dissimilar substrates by the *Escherichia coli* multidrug transporter MdfA. *Biochemistry* **40,** 12612–12618 (2001).
76. Loe, D. W., Deeley, R. G. & Cole, S. P. C. Characterisation of vincristine transport by the M_r 190,000 multidrug resistance protein (MRP): evidence for cotransport with reduced glutathione. *Cancer Res.* **58,** 5130–5136 (1998).
77. Rius, M., Mies, A. J., Hummel-Eisenbeiss, J., Jedlitschky, G. & Keppler, D. Cotransport of reduced glutathione with bile salts by MRP4 (ABCC4) localized to the basolateral hepatocyte membrane. *Hepatology* **38,** 374–384 (2003).

Acknowledgements This review is dedicated to Alex Neyfakh (1959–2006) who contributed much to our understanding of the multispecificity of drug binding.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Correspondence should be addressed to the author (chris.higgins@durham.ac.uk).

ARTICLES

Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia

Charles G. Mullighan[1]*, Salil Goorha[1]*, Ina Radtke[1], Christopher B. Miller[1], Elaine Coustan-Smith[2], James D. Dalton[1], Kevin Girtman[1], Susan Mathew[1]†, Jing Ma[5], Stanley B. Pounds[3], Xiaoping Su[5], Ching-Hon Pui[2], Mary V. Relling[4], William E. Evans[4], Sheila A. Shurtleff[1] & James R. Downing[1]

Chromosomal aberrations are a hallmark of acute lymphoblastic leukaemia (ALL) but alone fail to induce leukaemia. To identify cooperating oncogenic lesions, we performed a genome-wide analysis of leukaemic cells from 242 paediatric ALL patients using high-resolution, single-nucleotide polymorphism arrays and genomic DNA sequencing. Our analyses revealed deletion, amplification, point mutation and structural rearrangement in genes encoding principal regulators of B lymphocyte development and differentiation in 40% of B-progenitor ALL cases. The *PAX5* gene was the most frequent target of somatic mutation, being altered in 31.7% of cases. The identified *PAX5* mutations resulted in reduced levels of PAX5 protein or the generation of hypomorphic alleles. Deletions were also detected in *TCF3* (also known as *E2A*), *EBF1*, *LEF1*, *IKZF1* (*IKAROS*) and *IKZF3* (*AIOLOS*). These findings suggest that direct disruption of pathways controlling B-cell development and differentiation contributes to B-progenitor ALL pathogenesis. Moreover, these data demonstrate the power of high-resolution, genome-wide approaches to identify new molecular lesions in cancer.

Paediatric acute lymphoblastic leukaemia (ALL) comprises genetically distinct subtypes including B-progenitor leukaemias with translocations t(9;22)[*BCR–ABL1*], t(1;19)[*TCF3–PBX1*], t(12;21)[*ETV6–RUNX1*], rearrangements of *MLL*, hyperdiploid and hypodiploid karyotypes, and T-lineage leukaemia (T-ALL). These genetic lesions are important in leukaemia initiation[1], but alone are insufficient to generate a full leukaemic phenotype, indicating that cooperating oncogenic lesions are required. Although additional mutations have been identified in a subset of cases, the full complement of cooperating lesions and their distribution within the known genetic subtypes of ALL remain to be defined.

To obtain a comprehensive registry of genetic lesions in ALL, we examined DNA from the leukaemic cells (blasts) of 242 cases of paediatric ALL (Supplementary Table 1) using Affymetrix single nucleotide polymorphism (SNP) arrays that interrogate over 350,000 loci and permit identification of copy number changes at an average resolution of less than 5 kilobases (kb). In addition, paired copy number and loss of heterozygosity (LOH) analysis was performed for 228 ALL cases with matched remission samples (Supplementary Fig. 16 and Supplementary Table 15). Our analyses identified a mean of 6.46 somatic copy number alterations per case, with deletions outnumbering amplifications almost 2:1 (Table 1 and Supplementary Fig. 4). The frequency of genomic deletions and amplifications varied significantly between ALL subtypes (Table 1 and Supplementary Table 9). Genomic gains were frequent in B-progenitor ALLs with high hyperdiploidy, but were rare in other B-progenitor ALL subtypes. In contrast, deletions were more frequent, ranging from six deletions per case in some genetic subtypes (*ETV6–RUNX1* and hypodiploid ALL) to only a single deletion per case in *MLL*-rearranged ALL.

Fifty-four recurrent somatic regions of deletion were identified, with the minimal deleted regions typically measuring less than 1 megabase (Mb) in size, and twenty-four of the deletions containing only a single gene (Supplementary Table 10). None were present in the germline samples. Although technical aspects of the methodology used might theoretically lead to false-positive or false-negative results, fluorescence *in situ* hybridization (FISH) and/or quantitative polymerase chain reaction (qPCR) confirmatory studies (described below) validated each of the examined lesions. The recurring deletions included 3p14.2 (*FHIT*)[2], 6q16.2-3 (including *CCNC*)[3], 9p21.3 (two regions involving *CDKN2A* (ref. 4) and *MLLT3*), 12p13.2 (*ETV6*)[5], 11q23 (including *ATM*)[6], 13q14.2 (*RB1*)[7] and 13q14.2-3 (including *mir-16-1* and *mir-15a*)[8]. In addition, deletions of other tumour-associated genes not previously implicated in ALL were identified including *LEF1*, *BTG1* and *ERG*. The most notable observation, however, was the identification of genomic alterations in genes that regulate B-lymphocyte differentiation in 40% of B-progenitor ALL cases.

Focal deletions of *EBF1* in B-progenitor ALL

Eight B-progenitor ALL cases harboured mono-allelic deletions of *EBF1* (early B-cell factor), with deletions limited to this gene in six cases (Fig. 1a, b, Supplementary Figs 5–7 and Supplementary Table 11). EBF1 is required for the development of B cells, and with TCF3 regulates the expression of B-lineage-specific genes. Mice null for *Ebf1* arrest B-cell development at the pro-B-cell stage, whereas $Ebf1^{+/-}$ mice have a 50% reduction in the number of mature B cells but a normal number of pro-B cells[9]. These observations suggest that *EBF1* haploinsufficiency may contribute to leukaemogenesis. In support of this interpretation, FISH analysis of one case revealed two

[1]Departments of Pathology, [2]Oncology, [3]Biostatistics, [4]Pharmaceutical Sciences, and the [5]Hartwell Center for Bioinformatics and Biotechnology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. †Present address: The Department of Pathology & Laboratory Medicine, New York Presbyterian Hospital, Cornell Campus, 525 East 68th Street, F511, New York, New York 10021, USA.
*These authors contributed equally to this work.

Table 1 | Frequency of genomic amplifications and deletions in paediatric ALL

Group	Subtype	*N*	Amplifications (mean ± s.d.)*	Deletions (mean ± s.d.)*	All lesions (mean ± s.d.)*
B-ALL	Hyperdiploidy with >50 chromosomes	39	9.56 ± 3.59 (5–20)	1.59 ± 2.49 (0–11)	11.13 ± 5.0 (5–27)
B-ALL	*TCF3–PBX1*	17	1.59 ± 0.62 (1–3)	2.12 ± 1.17 (1–4)	3.7 ± 1.53 (2–7)
B-ALL	*ETV6–RUNX1*	47	0.89 ± 1.51 (0–8)	6.0 ± 4.63 (1–21)	6.68 ± 4.8 (0–21)
B-ALL	*MLL* rearranged	11	0.09 ± 0.3 (0–1)	0.91 ± 1.81 (0–6)	1 ± 1.79 (0–6)
B-ALL	*BCR–ABL1*	9	4 ± 5.3 (0–12)	4.2 ± 4.15 (0–12)	6.8 ± 4.52 (0–13)
B-ALL	Hyperdiploidy with 47–50 chromosomes	23	1.70 ± 1.55 (0–7)	3.5 ± 3.12 (0–12)	5.1 ± 4.31 (0–15)
B-ALL	Hypodiploid	10	1.1 ± 1.91 (0–6)	6.0 ± 4.42 (3–18)	7.1 ± 6.12 (3–24)
B-ALL	Other	36	1.06 ± 3.21 (0–19)	4.64 ± 5.14 (0–20)	5.58 ± 6.57 (0–23)
B-ALL	Total	192	2.97 ± 4.28 (0–20)	3.83 ± 4.2 (0–21)	6.63 ± 5.56 (0–27)
T-ALL		50	0.9 ± 1.98 (0–9)	4.9 ± 6.21 (0–30)	5.8 ± 7.12 (0–39)
All cases		242	2.54 ± 4.0 (0–20)	4.06 ± 4.69 (0–38)	6.46 ± 5.90 (0–39)

Lesions exclude inherited copy number variations[38], as well as focal deletions arising from antigen receptor gene rearrangements at 2p11.2 (*IGK@*), 7p14.1 (*TRG@*), 7q34 (*TRB@*), 14q11.2 (*TRAV*, *TRDV*, *TRDJ*, *TRDC* and *TRAJ*), 14q32.33 (*IGHV@*) and 22q11.22 (*IGL@*). Homozygous deletions are scored as two lesions. There is significant variation in the mean number of amplifications (analysis of variance, $P < 0.0001$), deletions ($P < 0.0001$) and all lesions ($P < 0.0001$) between B-progenitor ALL subtypes. Amplifications are more common in B-progenitor ALL than T-ALL (*t*-test, $P = 0.001$). Post-hoc pairwise comparisons of lesion frequency between B-progenitor ALL subtypes are shown in Supplementary Table 9.
* Range is shown in parentheses.

populations of blasts, one with mono-allelic and the other with bi-allelic *EBF1* deletions (Fig. 1c). Flow cytometric analysis of this case demonstrated two distinct blast populations: an immature pro-B-cell-like $CD10^{dim}CD22^{dim}$ population and a more mature $CD10^{bright}CD22^{bright}$ population (Fig. 1d). These populations significantly differed in their expression of CD79A (also called MB-1), a transcriptional target of EBF1 (Fig. 1e). Notably, FISH analysis on sorted blasts revealed homozygous deletion of *EBF1* in the immature $CD10^{dim}$ fraction and hemizygous *EBF1* deletions in the more mature $CD10^{bright}$ fraction (Fig. 1f and Supplementary Table 11). Thus, these data suggest that deletion of *EBF1* in these cases may contribute to the block in differentiation that is a hallmark of ALL.

In cases of ALL with mono-allelic deletion of *EBF1*, expression of the wild-type *EBF1* allele was detected by RT–PCR, consistent with haploinsufficiency (Supplementary Fig. 8). Moreover, no evidence of promoter methylation or point mutations of *EBF1* was detected (Supplementary Fig. 10)[10].

Frequent mono-allelic deletions of *PAX5*

We next examined the cohort for abnormalities in other genes within the B-cell development pathway (see Supplementary Table 12 and Supplementary Fig. 11). We identified copy number changes in *PAX5* in 57 out of 192 (29.7%) B-progenitor ALL cases (mono-allelic loss in 53, bi-allelic loss in 3, and an internal amplification in 1; Fig. 2 and Supplementary Figs 12–14). PAX5 is essential for B-lineage commitment and maintenance[11] and acts downstream of TCF3 and EBF1 to activate expression of B-lineage-specific genes such as *CD19*, *CD79A*, *BLNK* and *CD72*, and to repress expression of alternative lineage

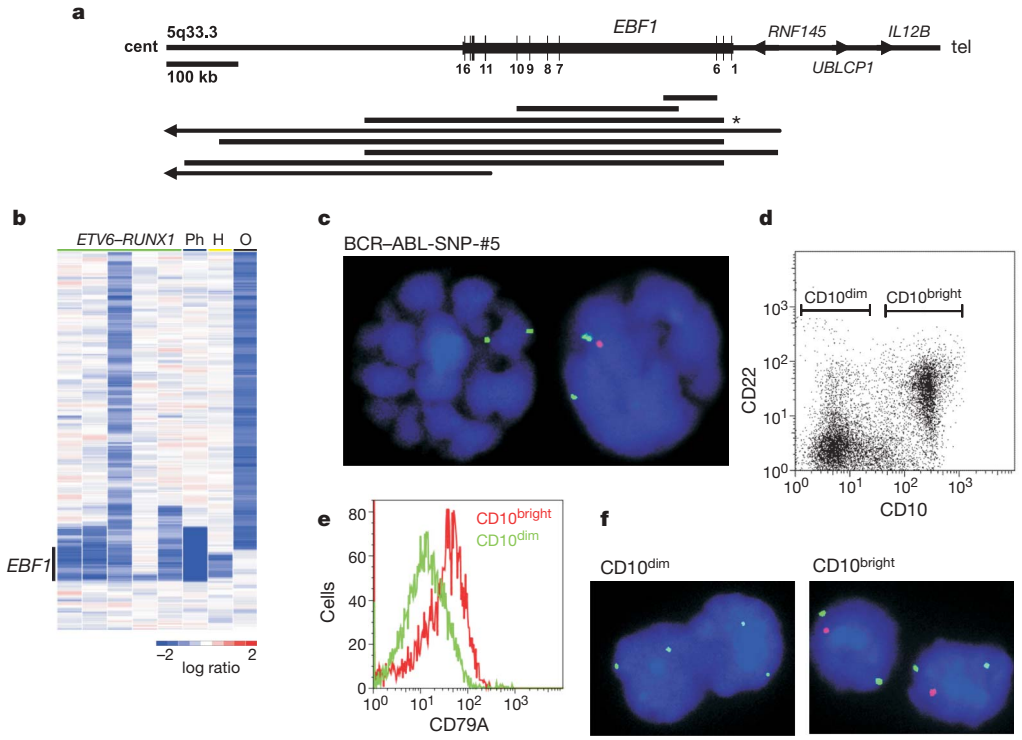


Figure 1 | *EBF1* deletions in B-progenitor ALL. a, *EBF1* exons are indicated by vertical lines. The extent of *EBF1* deletion in each case is shown by a horizontal line, with large deletions denoted by arrowed lines. Asterisk indicates case BCR–ABL-SNP#5 examined by FISH as shown in **c**. **b**, Copy number 'heat map' showing *EBF1* deletion in eight B-progenitor ALL cases. H, hypodiploid; O, other; Ph, *BCR–ABL1* positive. **c**, *EBF1* FISH analysis of case BCR–ABL-SNP#5 identifies two distinct blast populations, with either mono-allelic or bi-allelic *EBF1* deletion; the *EBF1* probe is red and the control probe is green. **d**, Two distinct blast populations differ in surface expression of CD22 and CD10. **e**, Correlation of blast CD79A and CD10 expression. **f**, FISH analysis of cells sorted by CD10 expression.

genes[12,13]. Four patterns of *PAX5* deletion were identified (Fig. 2b): (1) focal deletions involving only *PAX5* (25 cases); (2) broader deletions involving *PAX5* and a variable number of flanking genes (7 cases); (3) large 9p deletions involving the 3′ portion of *PAX5* (5 cases); and (4) deletion of all of chromosome 9 or 9p (19 cases). The *PAX5* deletions and amplification were confirmed by FISH and/or genomic qPCR (Supplementary Fig. 15 and Supplementary Tables 13, 14). FISH demonstrated that, with rare exception, the *PAX5* deletions were present in a dominant leukaemic clone, consistent with a role in leukaemogenesis.

Out of 25 B-progenitor ALLs with deletions confined to *PAX5*, 23 had deletions involving only a subset of *PAX5* exons, resulting in the expression of internally deleted transcripts (Supplementary Figs 17, 18 and Supplementary Table 16). In nine cases, in-frame splicing across the deletions would be predicted to encode an internally deleted PAX5 protein that lacks the DNA-binding domain and/or transcriptional regulatory domains. Western blot analysis in cases with sufficient material revealed expression of the predicted size mutant PAX5 protein (Fig. 2d, e and data not shown). The other transcripts are predicted to encode prematurely truncated proteins lacking key PAX5 functional domains. The focal amplification of *PAX5* exons 2–5 in case Hypodip-SNP-#7 is predicted to abolish expression of normal *PAX5* from the amplified allele. Thus, of the 57 cases with *PAX5* copy number changes, most lack expression of *PAX5* from the altered allele (27 cases), produce proteins that lack the DNA-binding domain (22 cases) or transcriptional regulatory domain (7 cases), or have a biallelic promoter mutation (one case), suggesting that these genetic alterations primarily result in loss of function.

Flow cytometry performed on leukaemic blasts in 17 cases demonstrated an approximately twofold reduction in the level of wild-type PAX5 protein in cases with mono-allelic deletions, indicative of haploinsufficiency (Supplementary Figs 19–21). Consistent with this interpretation, *PAX5* messenger RNA levels were quantified in 46 B-progenitor ALL cases (26 wild type; 20 mutated), and were found to be reduced in cases with *PAX5* deletions (Supplementary Fig. 22). A small number of *PAX5* wild-type cases also had low *PAX5* mRNA levels, raising the possibility of epigenetic *PAX5* silencing in these cases.

Cryptic translocations and point mutations of *PAX5*

Five B-progenitor ALL cases had large 9p deletions extending into the 3′ portion of *PAX5*. Two had rearrangements of chromosomes 9 and 12, either as a dicentric (9;12) or a t(9;12), and were shown by FISH and RT–PCR analyses to harbour the previously described *PAX5–ETV6* fusion[14,15] (Fig. 3a, b and Supplementary Figs 23a and 24). Rapid amplification of complementary DNA ends identified novel *PAX5* chimaeric transcripts in two of the remaining three cases: fusion of *PAX5* exon 6 to exon 7 of the forkhead box P1 gene (*FOXP1*), and fusion of *PAX5* exon 7 to exon 4 of the zinc finger protein 521 gene (*ZNF521*, the human homologue of murine *Zfp521* (also called *Evi3*)) (Fig. 3a). These were confirmed by FISH (Fig. 3c, d), RT–PCR (Supplementary Fig. 24) and sequencing analyses (Supplementary Fig. 23b, c). The *PAX5–ETV6* fusion was also detected as a 65-kDa protein in both cases containing this rearrangement (Fig. 2e); primary patient material was not available for western blot analysis of the *PAX5–FOXP1* and *PAX5–ZNF521* cases. A further 50

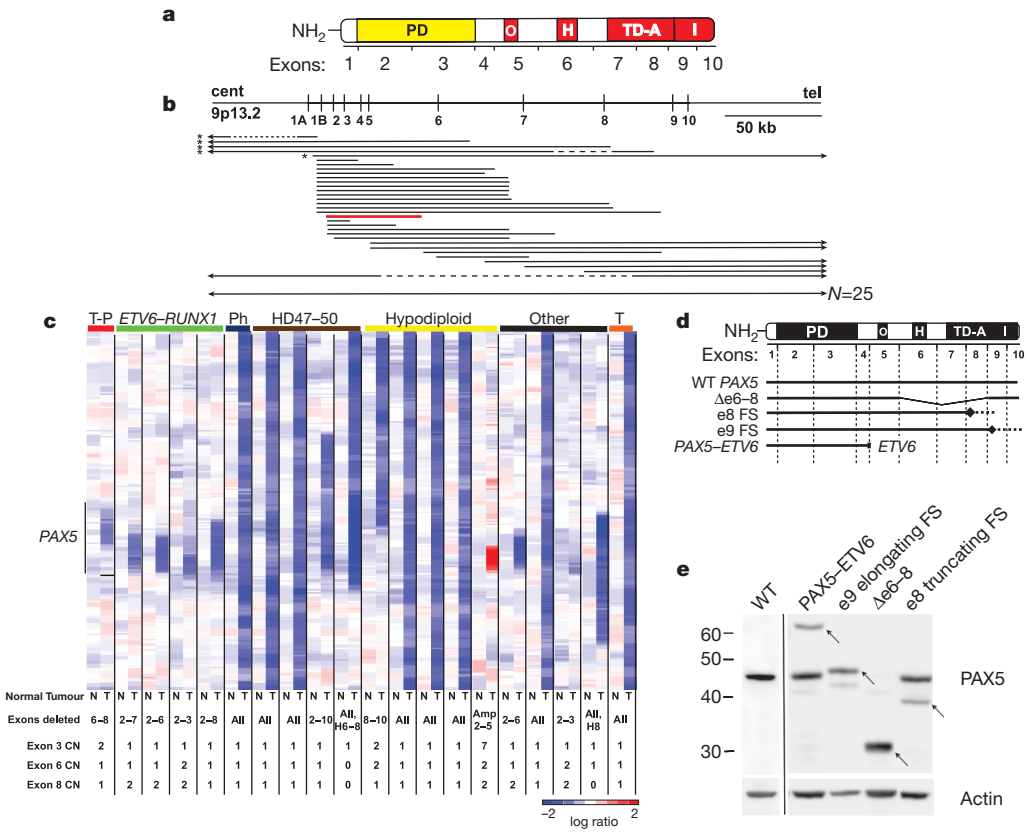


Figure 2 | *PAX5* deletions in ALL. **a**, *PAX5* exons and domains. H, homeodomain-like; I, transactivation, inhibitory; O, octapeptide; PD, paired domain; TD-A, transactivation, activating. **b**, Extent of *PAX5* deletions: hemizygous, solid lines; homozygous, dashed lines; amplification, red line. Deletions that stretch beyond the region shown are arrowed. The asterisk shows a deletion confined to *PAX5*. **c**, *PAX5* copy number heat map for 20 ALL cases and corresponding germline samples. *PAX5* exon-specific copy number as defined by quantitative genomic PCR is also shown. Amp, amplified; CN, copy number; H, homozygous; N, normal (remission sample); Ph, *BCR–ABL1*; T, T-ALL; T-P, *TCF3–PBX1*; H47–50, hyperdiploidy with 47–50 chromosomes. **d**, **e**, Schematic of *PAX5* mutations and western blot of primary leukaemic blasts. Mutant PAX5 proteins are indicated by arrows, and size in kDa is shown. FS, frameshift; WT, wild type.

cases were screened by RT–PCR for the three translocations and no additional cases were identified.

Each of these *PAX5* chimaeric genes encodes a protein that fuses the DNA-binding paired domain of PAX5 to the DNA-binding and transcriptional regulatory domains of the partner proteins (Fig. 3a). Thus, the fusion proteins are predicted to retain the ability to bind to PAX5 transcriptional targets, but would no longer provide normal transcriptional regulatory functions. The fusion proteins may also influence the expression of genes normally regulated by the partner protein, each of which has been implicated in B-cell development or haematopoietic malignancies[16–20].

Sequencing of *PAX5* identified point mutations in 14 cases that clustered in exons encoding the DNA-binding or transcriptional regulatory domains (Fig. 3e, Supplementary Fig. 25 and Supplementary Tables 17–19). In 13 cases with available germline DNA, the mutations were shown to be somatically acquired. Moreover, a quantitative analysis suggested that the mutations were present in all blasts in 12 out of 14 cases, and in a major subclone in the remaining two cases (Supplementary Table 18). The point mutations were hemizygous except in one case, where a homozygous splice-site mutation (IVS9+1) of *PAX5* was detected within an extensive region of 9p LOH that included a homozygous deletion of *CDKN2A* (Supplementary Fig. 16f and Supplementary Table 15).

Nine *PAX5* point mutations were identified within exons encoding the paired DNA-binding domain: V26G, P34Q, P80R and a complex insertion/deletion that replaces the amino acid sequence NDTVP at codons 126–130 with RA (Fig. 3e). Modelling studies using the PAX5 crystal structure[21] suggest that each point mutation should impair DNA binding (Supplementary Fig. 26). Other point mutations consist of frameshift, splice site, or missense mutations that affect the transactivation domain, and a single case with an exon 1B frameshift mutation that results in a prematurely truncated ten-residue polypeptide (Fig. 3e and Supplementary Table 17). Collectively, the identified *PAX5* point mutations are predicted to result in lost or altered DNA-binding or transcriptional regulatory function.

We also assessed whether methylation-induced silencing of *PAX5* occurs in B-progenitor ALL using mass spectrometry[10]. These data revealed high-level *PAX5* promoter methylation in T-ALL, but minimal methylation in B-progenitor ALL, irrespective of *PAX5* mutational status (Supplementary Figs 9, 10 and 27). Thus, epigenetic silencing owing to high-level methylation does not seem to be a prominent mechanism of *PAX5* silencing in B-progenitor ALL.

Functional consequences of *PAX5* mutations

To assess the DNA-binding and transcriptional activity of a subset of the identified *PAX5* mutants, luciferase-based reporter assays were performed using the PAX5-dependent reporter plasmid, *luc-CD19* (ref. 22). Each of the mutant PAX5 proteins tested displayed significantly reduced transcriptional activation compared to wild-type PAX5 (Fig. 4a, b and Supplementary Fig. 28). Moreover, transfection of increasing amounts of *PAX5–ETV6* or *PAX5–FOXP1* together with a fixed amount of wild-type *PAX5* demonstrated that the fusion proteins competitively inhibit the transcriptional activation of wild-type PAX5 (Supplementary Fig. 29 and Supplementary Table 20). Analysis of DNA-binding activity revealed a marked reduction in binding activity for PAX5 variants with paired domain mutations or deletions (Supplementary Fig. 30 and data not shown).

We next examined the effect of PAX5 mutations on transcriptional activation of the PAX5 target *Cd79a* (also called *mb-1* or *Ig-α*) in the murine plasmacytoma cell line 558LμM[23]. This cell line expresses three of the four components of the surface IgM (sIgM) receptor complex (μ heavy chain, λ light chain and Ig-β), but not Cd79a or Pax5, and consequently does not express sIgM. After transduction with *PAX5*-expressing retrovirus, Cd79a, and thus sIgM, are expressed, and quantification of sIgM serves as a sensitive measure of the *Cd79a* trans-activating activity of PAX5 (ref. 23). This assay

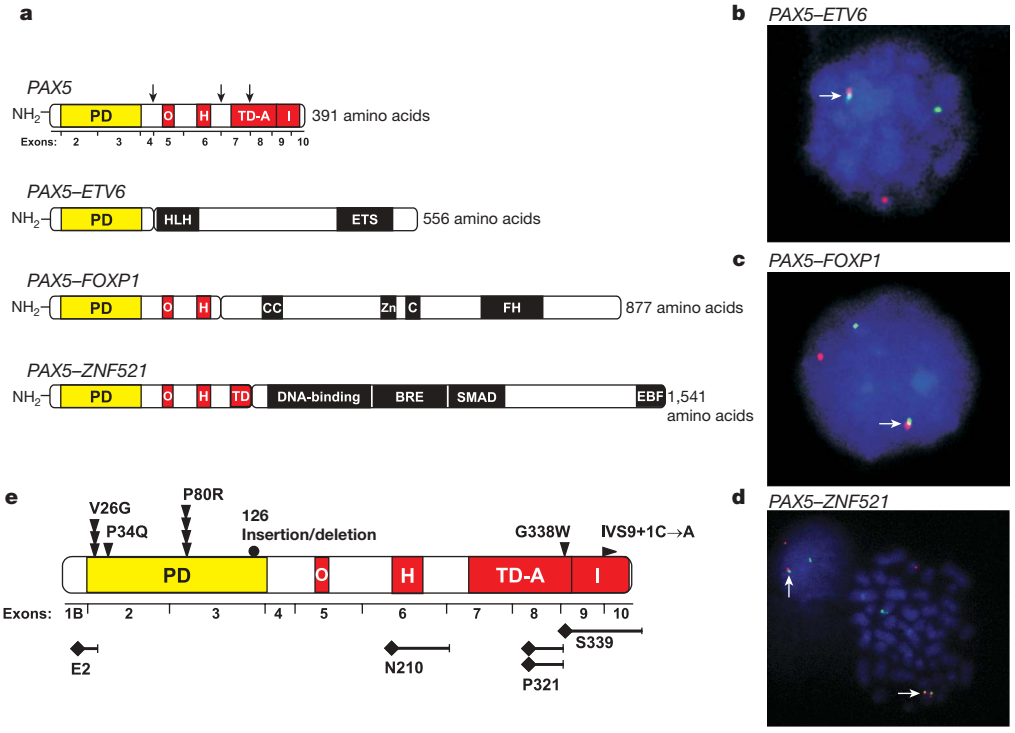


Figure 3 | *PAX5* translocations in B-progenitor ALL. a, Schematic of PAX5 translocation-encoded fusion proteins. Breakpoints are indicated by downward pointing arrows. BRE, BMP2 response element domain; C and CC, coiled-coil domain; EBF, EBF-interaction domain; ETS, Ets domain; FH, forkhead domain; H, homeodomain-like; HLH, helix–loop–helix; O, octapeptide domain; PD, paired domain; SMAD, SMAD-interacting domain; Zn, zinc finger domain. **b–d**, Interphase FISH analysis for *PAX5–ETV6* (**b**), *PAX5–FOXP1* (**c**) and *PAX5–ZNF521* (**d**). The *PAX5*-specific probe is red, and the partner genes green; arrows denote the fusions. **e**, Location of missense (downward pointing arrowheads), insertion/deletion (filled circles), frameshift (filled diamonds) and splice-site (right-pointing arrowhead) *PAX5* mutations.

confirmed the reduced transcriptional activity of the PAX5 mutants (Fig. 4c, d and Supplementary Fig. 31). Co-transduction with wild-type and mutant *PAX5*-expressing retroviruses showed that wild-type *PAX5* restored sIgM expression in the presence of the P34Q, P80R or Δ6–8 *PAX5* variants, but the restoration was incomplete in the presence of *PAX5–ETV6* (Fig. 4e and Supplementary Fig. 32). This suggests that the PAX5 DNA-binding and internal deletion mutants act as hypomorphic alleles with weak competitive activity, whereas *PAX5–ETV6* acts as a stronger competitive inhibitor of wild-type PAX5.

Because the *PAX5* mutations are predicted to reduce or inhibit normal PAX5 functional activity, we next examined expression of PAX5 target genes in leukaemic blasts. We did not observe a correlation between *PAX5* mutation status and expression of the PAX5 targets CD19 and CD79A[13,24]. However, using Affymetrix HG-U133A gene expression profiling data[25], we identified a 42-gene expression signature for *PAX5*-mutated *ETV6–RUNX1* B-progenitor ALL cases that included both upregulated (PAX5 repressed) and downregulated (PAX5 stimulated) genes (Supplementary Table 21 and Supplementary Fig. 33a). PAX5-stimulated genes included the known PAX5 target *CD72* (ref. 13) as well as genes with known roles in oncogenesis (for example, *DAPK1* (ref. 26), *TACC1* (ref. 27)). We then performed cross-subtype gene-set enrichment analysis[28] to test for enrichment of these putative PAX5-regulated genes in non-*ETV6–RUNX1* B-progenitor ALL cases. This analysis demonstrated highly significant enrichment of the PAX5-stimulated genes in *PAX5* wild-type B-progenitor ALL cases that lacked recurrent cytogenetic abnormalities, and in *PAX5* wild-type cases from the entire non-*ETV6-RUNX1* B-progenitor ALL cohort (Supplementary Fig. 33b, c). This suggests that the identified *PAX5* mutations have a major effect on the intracellular transcriptional network within primary leukaemic cells.

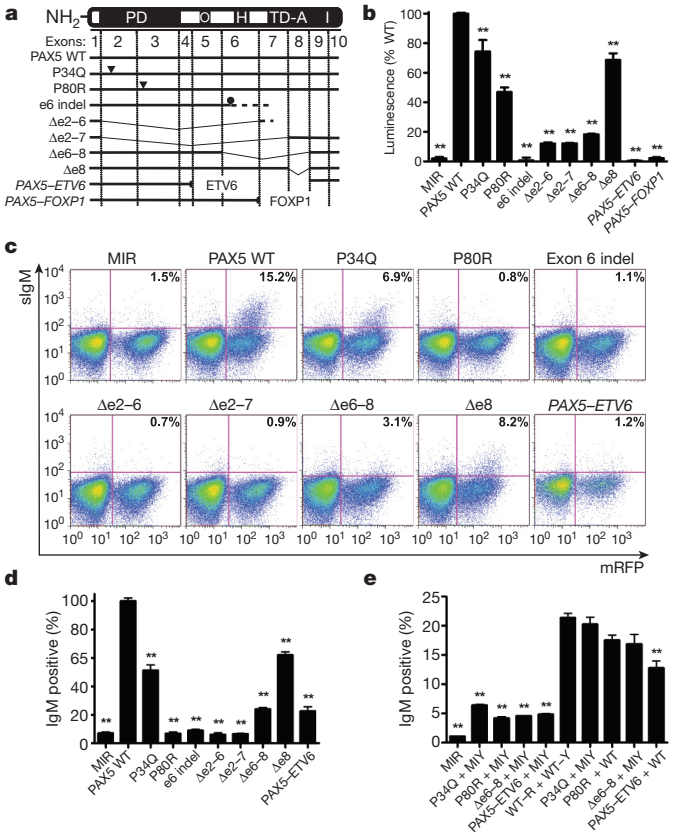


Figure 4 | Impaired function of PAX5 mutants. **a**, Structure of PAX5 variants. Indel, insertion/deletion. **b**, Transcriptional activity of PAX5 variants using the *luc*-CD19 reporter. Bars show mean (±s.e.m.) luciferase activity of triplicate experiments normalized to luciferase activity of wild-type (WT) PAX5. Double asterisk, analysis of variance with Dunnett's test, $P < 0.01$ compared with wild-type PAX5. MIR, MSCV-IRES-mRFP retrovirus not expressing PAX5. **c**, Representative flow plots of 558LμM cells transduced with *PAX5*-expressing retrovirus, showing reduced sIgM expression of PAX5 mutants compared with wild-type PAX5. Percentages indicate proportion of $mRFP^+$ cells that are $sIgM^+$. mRFP, expression of monomeric red fluorescent protein. **d**, Mean (±s.e.m.) 558LμM sIgM expression of triplicate experiments, normalized to sIgM expression of wild-type *PAX5*-transduced 558LμM cells. **e**, Impaired sIgM expression is corrected after co-infection with wild-type *PAX5* retrovirus (WT) for the PAX5 deletion and point mutants, but not *PAX5–ETV6*. Data show mean ±s.e.m.

Pattern of B-cell-development gene mutations in ALL

Deletions involving additional genes encoding regulators of B-cell development were identified, including *IKZF1* (17 B-progenitor ALL cases), *IKZF3* (3 cases), *LEF1* (3 cases) *TCF3* (1 case) and *BLNK* (2 cases; Supplementary Tables 22–24). Genomic sequencing of *IKZF1* was performed in all cases and no mutations were identified.

Of note, there were marked differences in the frequency and type of mutations among genetic subtypes of ALL (Supplementary Table 22–24). Specifically, all ten hypodiploid cases had one null *PAX5* allele, and six of these also had either a point mutation (five cases) or translocation (*PAX5–ZNF521*, one case) involving the other *PAX5* allele. Five of the ten hypodiploid cases also contained deletions in other B-cell-development genes, with up to three different genes being mutated within a single case. In contrast, 28% of the *ETV6–RUNX1* cases contained focal mono-allelic *PAX5* deletions but lacked evidence of mutations in the retained *PAX5* allele. Three of these cases also had deletions of a single allele of *EBF1*, and another two cases had deletions of only *EBF1*. At the other end of the spectrum, mutations in B-cell-development genes were uncommon in hyperdiploid B-progenitor ALL cases (13% of cases).

Discussion

Understanding the molecular pathogenesis of cancer requires a detailed cataloguing of all genetic lesions within a cancer cell. The data presented in this paper provide a rational 'road map' for approaching such a task. We have shown that genomic copy number analysis using high-density SNP arrays can pinpoint altered genes and pathways for further analysis. This approach led to the unexpected finding of mutations in genes encoding regulators of B-cell development and differentiation in 40% of B-progenitor ALL cases. The identified genetic alterations are specific, pathogenic and somatically acquired. In addition, most *EBF1* and *PAX5* deletions affect only these genes, thus conclusively identifying them as the target of the deletions. Moreover, the average number of deletions within an individual B-progenitor ALL case was 3.83, and focal amplifications were uncommon, indicating that global genomic instability is not an underlying mechanism. Finally, although somatic point mutations were identified in *PAX5*, sequencing of *EBF1* and *IKZF1* failed to reveal any evidence of mutations, ruling out a high mutational rate.

The most common targets of these genetic alterations (*EBF1*, *PAX5* and *IKZF1*) have central roles in the development of normal B cells[12]. In mice, the complete absence of *Ebf1* or *Pax5* results in the arrest of B-cell development at the early pro-B- or pro-B-cell stage[9,11], and loss of *Ikzf1* leads to an arrest at an even earlier stage of lymphoid development[29]. Moreover, haploinsufficiency of *Ebf1* leads to a partial block in B-cell development[9], a phenotype that is further accentuated in mice haploinsufficient for both *Ebf1* and *E2a* (also called *Tcfe2a*)[30]. Furthermore, loss of *Ikzf1* through the expression of a dominant-negative *Ikzf1* isoform predisposes mice to the development of T-lineage malignancies[31].

The overall consequence of the identified lesions is to reduce the level of the specific transcription factor either as a result of mono-allelic deletion or the generation of altered forms of the specific protein. Whether some mutations result in dominant-negative forms

of PAX5, or alternatively alter transcriptional activity in a promoter specific-manner, requires further investigation. It is important to note that during the normal development of B cells, *PAX5* is subjected to allele-specific regulation, with only a single *PAX5* allele transcribed during the earliest phase of B-cell commitment and then a switch to bi-allelic expression as B cells begin to differentiate[32]. The loss of a wild-type *PAX5* allele from the identified mutations would eliminate the ability to turn on normal bi-allelic transcription, which may directly contribute to the differentiation arrest seen in ALL.

Members of the PAX family of transcription factors, including *PAX3*, *PAX7* and *PAX8*, have previously been identified as targets of tumour-associated translocations[33] that result in the overexpression of PAX fusion proteins. Furthermore, overexpression of wild-type *PAX5* occurs as a result of its rearrangement into the *IGH* locus in non-Hodgkin's lymphoma[34]. In contrast to overexpression, however, our results implicate subtle changes in the dosage of PAX5 and other key regulators of B-cell development in the pathogenesis of ALL. Although heterozygous alterations of *PAX5* may have only subtle effects on their own, these effects probably contribute directly to leukaemogenesis in combination with other oncogenic lesions.

The identified high frequency of alterations of B-cell-development genes in paediatric B-progenitor ALL represents a lower limit of the true frequency. Direct copy number analysis of those genes with low-density SNP coverage along with full sequence analysis of all genes involved in controlling B-cell development and differentiation will be required to define the overall frequency. Nevertheless, our results demonstrate an unexpectedly high frequency of alterations with distinct patterns of mutation among the various genetic subtypes of paediatric ALL. Experiments to assess directly the effect of co-expressing *PAX5* mutants and fusion proteins such as *ETV6–RUNX1* in murine models should provide valuable insights into the ability of these lesions to collaborate in leukaemogenesis. Moreover, attempts to determine whether small-molecule inducers of differentiation can by-pass the block resulting from the identified lesions, and whether these molecules, in turn, would trigger a leukaemia-cell-specific apoptotic response, could lead to new therapeutic approaches for paediatric B-progenitor ALL.

METHODS

Two-hundred-and-forty-two B- and T-lineage paediatric ALL cases treated at St Jude Children's Research Hospital (SJCRH) were studied (Supplementary Table 1). DNA extracted from leukaemic blasts and remission marrow or blood samples was genotyped with Affymetrix GeneChip Human Mapping 50K Hind 240, 50K Xba 240 and 250K Sty arrays (Affymetrix)[35]. DNA copy number and paired LOH analyses were performed using dChipSNP[36] and circular binary segmentation[37]. To improve the accuracy of copy number inference by both algorithms, we developed a normalization procedure that uses SNPs exclusively from regions shown to be diploid in the leukaemic blasts by routine cytogenetics (Supplementary Fig. 2). Bioinformatic and biostatistical analyses, FISH assays, flow cytometry and cell sorting, genomic sequencing and methylation analysis of *EBF*, *PAX5* and *IKZF1*, qPCR, *PAX5* variant and fusion cloning, luciferase reporter and gel-shift assays, 558LμM assays, western blotting, and cross-subtype gene-set enrichment analyses are detailed in the Supplementary Methods.

Received 7 November 2006; accepted 20 February 2007.
Published online 7 March 2007.

1. Greaves, M. F. & Wiemels, J. Origins of chromosome translocations in childhood leukaemia. *Nature Rev. Cancer* **3,** 639–649 (2003).
2. Peters, U. R. *et al.* Aberrant FHIT mRNA transcripts are present in malignant and normal haematopoiesis, but absence of FHIT protein is restricted to leukaemia. *Oncogene* **18,** 79–85 (1999).
3. Jackson, A. *et al.* Deletion of 6q16-q21 in human lymphoid malignancies: a mapping and deletion analysis. *Cancer Res.* **60,** 2775–2779 (2000).
4. Okuda, T. *et al.* Frequent deletion of p16INK4a/MTS1 and p15INK4b/MTS2 in pediatric acute lymphoblastic leukemia. *Blood* **85,** 2321–2330 (1995).
5. Raynaud, S. *et al.* The 12;21 translocation involving TEL and deletion of the other TEL allele: two frequently associated alterations found in childhood acute lymphoblastic leukemia. *Blood* **87,** 2891–2899 (1996).
6. Raimondi, S. C. *et al.* Acute lymphoblastic leukemias with deletion of 11q23 or a novel inversion (11)(p13q23) lack MLL gene rearrangements and have favorable clinical features. *Blood* **86,** 1881–1886 (1995).
7. Heerema, N. A. *et al.* Abnormalities of chromosome bands 13q12 to 13q14 in childhood acute lymphoblastic leukemia. *J. Clin. Oncol.* **18,** 3837–3844 (2000).
8. Cave, H. *et al.* Deletion of chromosomal region 13q14.3 in childhood acute lymphoblastic leukemia. *Leukemia* **15,** 371–376 (2001).
9. Lin, H. & Grosschedl, R. Failure of B-cell differentiation in mice lacking the transcription factor EBF. *Nature* **376,** 263–267 (1995).
10. Ehrich, M. *et al.* Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc. Natl Acad. Sci. USA* **102,** 15785–15790 (2005).
11. Nutt, S. L., Heavey, B., Rolink, A. G. & Busslinger, M. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* **401,** 556–562 (1999).
12. Busslinger, M. Transcriptional control of early B cell development. *Annu. Rev. Immunol.* **22,** 55–79 (2004).
13. Ying, H., Healy, J. I., Goodnow, C. C. & Parnes, J. R. Regulation of mouse CD72 gene expression during B lymphocyte development. *J. Immunol.* **161,** 4760–4767 (1998).
14. Cazzaniga, G. *et al.* The paired box domain gene *PAX5* is fused to *ETV6/TEL* in an acute lymphoblastic leukemia case. *Cancer Res.* **61,** 4666–4670 (2001).
15. Strehl, S., Konig, M., Dworzak, M. N., Kalwak, K. & Haas, O. A. *PAX5/ETV6* fusion defines cytogenetic entity dic(9;12)(p13;p13). *Leukemia* **17,** 1121–1123 (2003).
16. Bohlander, S. K. ETV6: a versatile player in leukemogenesis. *Semin. Cancer Biol.* **15,** 162–174 (2005).
17. Wlodarska, I. *et al. FOXP1*, a gene highly expressed in a subset of diffuse large B-cell lymphoma, is recurrently targeted by genomic aberrations. *Leukemia* **19,** 1299–1305 (2005).
18. Bond, H. M. *et al.* Early hematopoietic zinc finger protein (EHZF), the human homolog to mouse Evi3, is highly expressed in primitive human hematopoietic cells. *Blood* **103,** 2062–2070 (2004).
19. Hu, H. *et al.* Foxp1 is an essential transcriptional regulator of B cell development. *Nature Immunol.* **7,** 819–826 (2006).
20. Warming, S. *et al. Evi3*, a common retroviral integration site in murine B-cell lymphoma, encodes an EBFAZ-related Kruppel-like zinc finger protein. *Blood* **101,** 1934–1940 (2003).
21. Garvie, C. W., Hagman, J. & Wolberger, C. Structural studies of Ets-1/Pax5 complex formation on DNA. *Mol. Cell* **8,** 1267–1276 (2001).
22. Czerny, T. & Busslinger, M. DNA-binding and transactivation properties of Pax-6: three amino acids in the paired domain are responsible for the different sequence recognition of Pax-6 and BSAP (Pax-5). *Mol. Cell. Biol.* **15,** 2858–2871 (1995).
23. Maier, H., Colbert, J., Fitzsimmons, D., Clark, D. R. & Hagman, J. Activation of the early B-cell-specific *mb-1* (Ig-alpha) gene by Pax-5 is dependent on an unmethylated Ets binding site. *Mol. Cell. Biol.* **23,** 1946–1960 (2003).
24. Nutt, S. L., Morrison, A. M., Dorfler, P., Rolink, A. & Busslinger, M. Identification of BSAP (Pax-5) target genes in early B-cell development by loss- and gain-of-function experiments. *EMBO J.* **17,** 2319–2333 (1998).
25. Ross, M. E. *et al.* Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood* **102,** 2951–2959 (2003).
26. Martoriati, A. *et al. dapk1*, encoding an activator of a $p19^{ARF}$-p53-mediated apoptotic checkpoint, is a transcription target of p53. *Oncogene* **24,** 1461–1466 (2005).
27. Conte, N. *et al.* Carcinogenesis and translational controls: TACC1 is down-regulated in human cancers and associates with mRNA regulators. *Oncogene* **21,** 5619–5630 (2002).
28. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102,** 15545–15550 (2005).
29. Georgopoulos, K. *et al.* The Ikaros gene is required for the development of all lymphoid lineages. *Cell* **79,** 143–156 (1994).
30. O'Riordan, M. & Grosschedl, R. Coordinate regulation of B cell differentiation by the transcription factors EBF and E2A. *Immunity* **11,** 21–31 (1999).
31. Winandy, S., Wu, P. & Georgopoulos, K. A dominant mutation in the *Ikaros* gene leads to rapid development of leukemia and lymphoma. *Cell* **83,** 289–299 (1995).
32. Nutt, S. L. *et al.* Independent regulation of the two *Pax5* alleles during B-cell development. *Nature Genet.* **21,** 390–395 (1999).
33. Robson, E. J., He, S. J. & Eccles, M. R. A. PANorama of PAX genes in cancer and development. *Nature Rev. Cancer* **6,** 52–62 (2006).
34. Busslinger, M., Klix, N., Pfeffer, P., Graninger, P. G. & Kozmik, Z. Deregulation of *PAX-5* by translocation of the Eμ enhancer of the *IgH* locus adjacent to two alternative *PAX-5* promoters in a diffuse large-cell lymphoma. *Proc. Natl Acad. Sci. USA* **93,** 6129–6134 (1996).
35. Matsuzaki, H. *et al.* Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nature Methods* **1,** 109–111 (2004).
36. Lin, M. *et al.* dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* **20,** 1233–1240 (2004).

37. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5,** 557–572 (2004).
38. McCarroll, S. A. *et al.* Common deletion polymorphisms in the human genome. *Nature Genet.* **38,** 86–92 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Z. Cai and L. King for technical assistance; B. Schulman for modelling of *PAX5* mutations; M. Roussel, M. Busslinger and J. Hagman for providing reagents; the Tumor Processing Laboratory of SJCRH for providing tumour samples; and C. Li for discussions and modifications of dChipSNP. This work was supported by grants from the National Cancer Institute (to J.R.D. and W.E.E.), the National Institute of General Medical Sciences (to M.V.R.), the National Health and Medical Research Council (Australia) (to C.G.M.), the Royal Australasian College of Physicians (to C.G.M.), the Haematology Society of Australia and New Zealand (to C.G.M), and the American Lebanese and Syrian Associated Charities (ALSAC) of SJCRH.

Author Contributions J.R.D. designed and supervised experiments. C.G.M, S.G., I.R, C.B.M., E.C.-S., J.D.D and K.G. performed experiments. S.A.S. collected and managed clinical samples and data. C.G.M., S.G., J.M., S.B.P. and X.S. analysed data. S.B.P. developed software for DNA copy number analysis. C.-H.P., M.V.R. and W.E.E. provided patient samples. C.G.M. and J.R.D. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information The primary SNP microarray data have been deposited in NCBIs Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE5511. The data are also available at http://www.stjuderesearch.org/data/ALL-SNP1/. Sequences of the *PAX5* fusion transcripts have been deposited in GenBank with accessions DQ841178 (*PAX5–ETV6*), DQ845346 (PAX5–FOXP1) and DQ845345 (*PAX5–ZNF521*). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.R.D. (james.downing@stjude.org).

ARTICLES

Mediators of vascular remodelling co-opted for sequential steps in lung metastasis

Gaorav P. Gupta[1]*, Don X. Nguyen[1]*, Anne C. Chiang[1,2], Paula D. Bos[1], Juliet Y. Kim[1], Cristina Nadal[1]†, Roger R. Gomis[1]†, Katia Manova-Todorova[3] & Joan Massagué[1,4]

Metastasis entails numerous biological functions that collectively enable cancerous cells from a primary site to disseminate and overtake distant organs. Using genetic and pharmacological approaches, we show that the epidermal growth factor receptor ligand epiregulin, the cyclooxygenase COX2, and the matrix metalloproteinases 1 and 2, when expressed in human breast cancer cells, collectively facilitate the assembly of new tumour blood vessels, the release of tumour cells into the circulation, and the breaching of lung capillaries by circulating tumour cells to seed pulmonary metastasis. These findings reveal how aggressive primary tumorigenic functions can be mechanistically coupled to greater lung metastatic potential, and how such biological activities may be therapeutically targeted with specific drug combinations.

The emergence of disseminated metastases remains the primary cause of mortality in cancer patients[1–3]. Recent searches for genetic determinants of metastasis have led to the identification of gene sets, or 'signatures', for which the expression in primary tumours is associated with high risk of metastasis and poor disease survival[4–10]. It has been proposed that the expression of such genes in primary tumours might directly predispose cancer cells for growth in distant organs[11], raising questions about the distinction between tumorigenic genes and genes that mediate metastasis. A contrasting, long-held view is that metastasis arises from rare tumour cell clones, the genetic make-up of which endows them with a unique selective advantage in distant organ microenvironments[1]. Breast cancer metastasis genes that may correspond to each of these two models have been recently identified in a functional screen for mediators of lung colonization[12]. A subset of the genes identified in this manner supports mammary tumour growth as well as pulmonary metastasis by human cancer cells in mice, and constitutes a lung metastasis gene signature (LMS), the expression of which in primary tumours indicates a high risk of pulmonary relapse in breast cancer patients[12,13]. Other genes emerging from the same screen, however, imparted lung metastasis virulence without affecting primary tumour growth[12]. The specific roles that these various genes may have in metastasis have remained unknown.

Here we report that four LMS genes collectively contribute vascular remodelling functions that can support the formation of vasculature in mammary tumours, the entry of tumour cells into the circulation and the exit of tumour cells from the bloodstream into the lung parenchyma. Their ability to mediate distinct functions in the primary site and in the lung metastasis setting distinguishes these genes from oncogenes that primarily support the transformed state in cancer cells. The present findings suggest a molecular basis for the lung metastasis proclivity of locally aggressive primary breast tumours, and a rationale for combinatorial therapeutic interventions against metastasis.

Genetic cooperation in mammary tumour growth

The MDA-MB-231 cell line, which was derived from the pleural fluid of a patient with widely metastatic breast cancer[14], is a heterogeneous population composed of cells with diverse organotropic metastatic potential and distinct pro-metastatic gene expression signatures[15–17]. Notably, some of the LMS genes that typify the lung metastatic subpopulations derived from this source were independently identified as downstream effectors of vascular endothelial growth factor (VEGF) in endothelial cells[18]. This suggested to us the possibility that, when expressed by tumour cells, these genes may confer vascular remodelling functions that are relevant for metastatic progression. These genes include the epidermal growth factor receptor (EGFR)/pan-HER ligand epiregulin (*EREG*), the prostaglandin-synthesizing enzyme cyclooxygenase 2 (*COX2*; also called *PTGS2*), and the matrix-remodelling metalloproteinases *MMP1* and *MMP2*. *EREG*, *COX2* and *MMP1* are part of the clinically validated LMS genes and *MMP2* is frequently associated with them[12]. Unlike endothelial cells[18], the lung metastatic MDA-MB-231 subpopulation LM2-4175 (hereafter called LM2) expressed these four genes in a manner that was neither dependent on autocrine VEGF (Supplementary Fig. 1a, b) nor responsive to VEGF addition (Supplementary Fig. 1c). Thus, we investigated whether the elevated expression of these genes in cancer cells might recapitulate a transcriptional programme encoding secretory mediators of vascular remodelling for tumorigenesis and lung metastasis.

We stably reduced the expression of *EREG*, *COX2*, *MMP1* and *MMP2* by using short hairpin RNA interference (shRNA) in LM2 cells[12]. We also generated compound knockdown cells in which up to four shRNAs were expressed in a given cell population, attaining significant silencing of all targeted genes and their encoded products (Fig. 1a and Supplementary Fig. 2). The extent of silencing achieved is consistent with multiple constructs being simultaneously active in most, but not all, cells in the knockdown population. Reduction of *EREG*, *COX2*, *MMP1* or *MMP2* expression individually had statistically significant, yet limited, effects on tumour growth on cell inoculation into the mammary fat pad of immunocompromised mice (Fig. 1b, left graph). Nevertheless, silencing of these genes in different combinations delayed tumour progression, with nearly complete abrogation of growth achieved by silencing all four genes simultaneously (Fig. 1b, right graph). Each cell line was infected with similar

[1]Cancer Biology and Genetics Program, [2]Department of Medicine, [3]Molecular Cytology Core Facility, and [4]Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. †Present addresses: Hemato-Oncology Institute, Hospital Clinic de Barcelona, 08036 Barcelona, Spain (C.N.); Oncology Programme, Institute for Research in Biomedecine, Barcelona Science Park and University of Barcelona, 08028 Barcelona, Spain (R.R.G.).
*These authors contributed equally to this work.

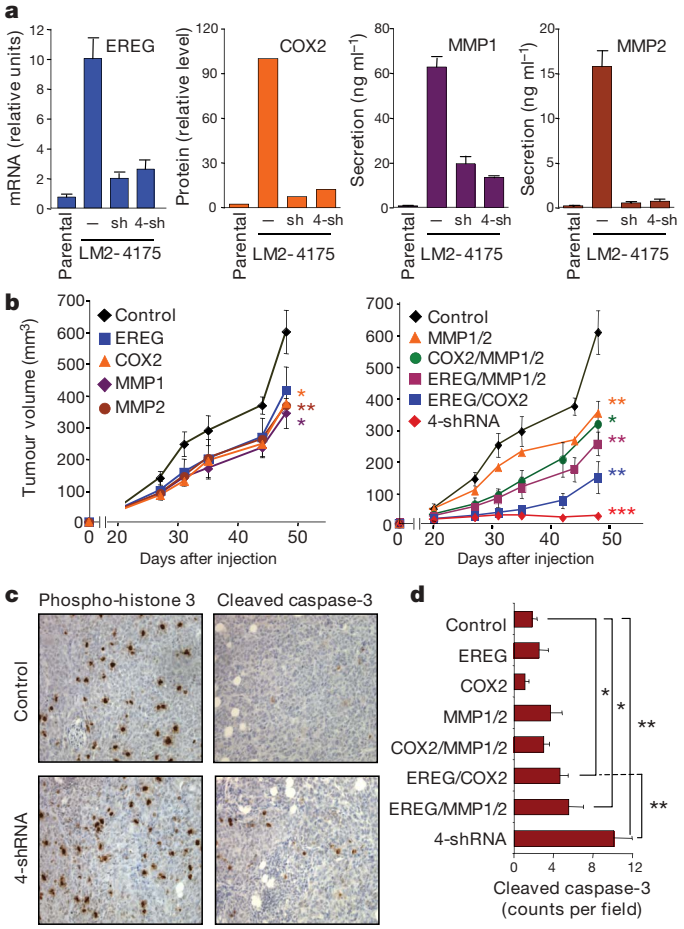


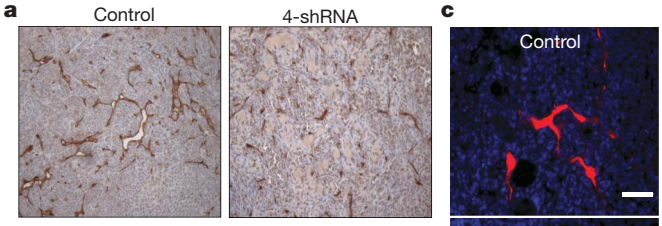
Figure 1 | *EREG*, *MMP1*, *MMP2* and *COX2* cooperate to mediate primary tumour growth. **a**, LM2 cells were infected with retrovirus encoding a control hairpin, or with shRNAs targeting *EREG*, *MMP1*, *MMP2* or *COX2*. For combination knockdown retrovirus, multiple hairpin vectors were transfected as pools into viral packaging cell lines. Infected cells were selected and *EREG* knockdown was determined by quantitative (q)RT–PCR, COX2 analysed via western blot, and secreted MMP1 and MMP2 measured by ELISA. Shown are levels of each gene product in the parental MDA-MB-231 cell line from which LM2 cells were selected, as well as LM2 control, single (sh) and quadruple knockdown (4-sh) cells. $n = 3$; error bars represent 95% confidence interval for qRT–PCR analysis and standard errors of the mean (s.e.m.) for ELISA. **b**, 1×10^6 cells of control, single knockdowns, or the indicated combination knockdown samples were inoculated into the fourth mammary fat pads of immunodeficient mice. Length and width of palpable tumours were measured, and tumour volumes calculated at the indicated time points. Left: effects of single gene knockdown; right: control compared to combination knockdown cells. $n = 6$; error bars indicate s.e.m.; asterisk, $P < 0.05$; double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$; calculated using a two-tailed Student's *t*-test for tumour volumes at the last time point, compared to control. **c**, Automated immunohistochemistry for phospho-histone 3 and cleaved caspase-3 detection was performed on tumours obtained from the various combination knockdown cell lines. Shown are representative images at an original magnification of ×20. **d**, Quantification of cleaved caspase-3 staining using Image J software. $n = 15$; error bars indicate s.e.m.; single asterisk, $P < 0.01$; double asterisk, $P < 0.001$; calculated using a two-sided Wilcoxon rank-sum test, compared to levels in control tumours.

retroviral titres irrespective of the number of different hairpin sequences introduced. Moreover, overexpression of MMP1, MMP2 and COX2 in the context of cells targeted by all four shRNAs resulted in phenotypic rescue of tumour growth to levels that were observed for the single *EREG* knockdown (Supplementary Fig. 3a, b). The combined knockdown of three lung metastasis virulence genes, *IL13RA2*, *SPARC* and *VCAM1* (ref. 12), only mildly affected mammary tumour growth despite strongly inhibiting lung metastasis (Supplementary Fig. 3d–f). Thus, not all combinations of lung metastasis mediators stimulate primary tumorigenesis. Overall, these results uncover specific genetic interactions between *EREG*, *MMP1*, *MMP2* and *COX2* that collectively facilitate accelerated mammary tumour growth.

Quantified phosphorylated histone H3 levels indicated that none of the single or combined knockdowns of these genes significantly altered the proliferation rate of the tumour cells (Fig. 1c and Supplementary Fig. 4a). In contrast, an increased rate of apoptosis was evident in tumours with combinatorial knockdown, as measured by cleaved caspase-3 staining (Fig. 1c, d). Dual inhibition of *EREG* and *COX2* resulted in a synergistic rise in the rate of primary tumour apoptosis (Supplementary Fig. 4b). Although further inhibition of *MMP1* and *MMP2* did not reach the statistical threshold for synergy, reduction of these genes resulted in a supra-additive effect above the level of apoptosis observed in the *EREG*/*COX2* knockdown (Supplementary Fig. 4c).

Role in mammary tumour angiogenesis

Increased rates of tumour cell death might be secondary to defects in angiogenesis. Indeed, histological staining for endothelial cell marker CD31 revealed profound defects in the vascular morphology of tumours that had reduced expression of all four of these lung metastasis genes (Fig. 2a). Digital imaging analysis and quantification of vessel structure demonstrated that although the average number of discrete vessel units was not considerably altered, the length, number of lumens and extent of branching of the existing vasculature were significantly reduced in the combined knockdown tumours (Fig. 2b). These morphological changes were also visualized using CD31 immunofluorescence and confocal microscopy to image the tumour vasculature (Supplementary Fig. 5). Co-staining for CD31 and NG2, a smooth muscle pericyte marker, did not reveal any major differences in pericyte recruitment (Supplementary Fig. 5b). Nevertheless, vessels in the quadruple knockdown tumours exhibited diminished effusion of intravenously injected dextran, consistent with attenuated vascular permeability (Fig. 2c). Of note, these defects in primary tumour vessel morphology and function occurred in the absence of differences in VEGF levels between control and knockdown tumour cells (Supplementary Fig. 1a). This suggested that the aforementioned genes promote the formation of the dilated, tortuous and



b

	Control	4-shRNA	
No. of vessel units (>10 μm)	75.7 ± 11.0	68.6 ± 4.4	
Average length (μm)	61.0 ± 8.6	37.6 ± 0.6	**
No. of lumen per field	43.5 ± 7.2	3.7 ± 0.6	***
No. of branch points per vessel	3.40 ± 0.75	1.26 ± 0.04	**

Figure 2 | *EREG*, *MMP1*, *MMP2* and *COX2* mediate tumour angiogenic progression. **a**, Automated anti-CD31 immunohistochemistry was performed on control and 4-shRNA-targeted tumours. Representative images were obtained at ×10 magnification. **b**, Tumour sections stained with anti-CD31 antibody were used for morphometric vessel analysis. Vessel units (defined as ≥10 μm in length), vessel length, number of lumen and average number of branch points were quantified as described in Methods. $n = 15$; error bars indicate s.e.m.; double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$; calculated using a two-tailed Student's *t*-test. **c**, Mice bearing tumours were injected with rhodamine-conjugated dextran. Tumours were extracted and sections examined for vessel permeability at ×10 magnification. Scale bars, 50 μm.

leaky blood vessels that typify the neovasculature of aggressive primary tumours[19].

Tumour cell extravasation from lung capillaries

To assess the importance of these genes as mediators of pulmonary colonization, the knockdown lines were injected intravenously into mice, and lung metastatic progression was monitored by bioluminescence (Fig. 3a) and histological examination (Supplementary Fig. 6a). Independent silencing of these genes had little impact in this lung colonization assay. As in the mammary tumorigenesis assays, inhibition of lung colonization was attained when these genes were silenced in combination, with knockdown of all four genes yielding the most salient defect. Metastatic colonies were eventually observed in the combination knockdown samples, but this was attributable to outgrowth of cells escaping knockdown and re-expressing these genes (data not shown). Statistical tests identified multiple synergistic interactions between *EREG*, *COX2*, *MMP1* and *MMP2* in the early progression of lung metastasis (Supplementary Fig. 6b). The specificity of this phenotype was confirmed by overexpressing MMP1, MMP2 and COX2, which resulted in significant recovery of lung metastatic activity (Supplementary Fig. 3c).

To visualize lung metastasis events during the initial days after inoculation, whole lungs were extracted and scanned by confocal microscopy. Within 2 days of entering the circulation, control tumour cells could be visualized outside of the lung capillaries, showing that they efficiently extravasate into the lung parenchyma (Fig. 3b, left panel). Conversely, when detected, the knockdown cells in which all four genes were targeted by shRNAs were trapped within vessels as single cells, seemingly incapable of breaching the lung endothelium (Fig. 3b, right panel). In an *in vitro* assay of trans-endothelial migration, the migratory capacity of LM2 cells through an endothelial monolayer was inhibited by the combined knockdown of all four genes, which did not entail a generalized defect in cell motility (Fig. 3c and Supplementary Fig. 7). Thus, although several mechanisms and outcomes have been proposed for interactions between EGFR/HER, MMPs and COX2[20–23], our results provide evidence that the expression of EREG, COX2, MMP1 and MMP2 by cancer cells can collectively promote metastatic extravasation in the lungs.

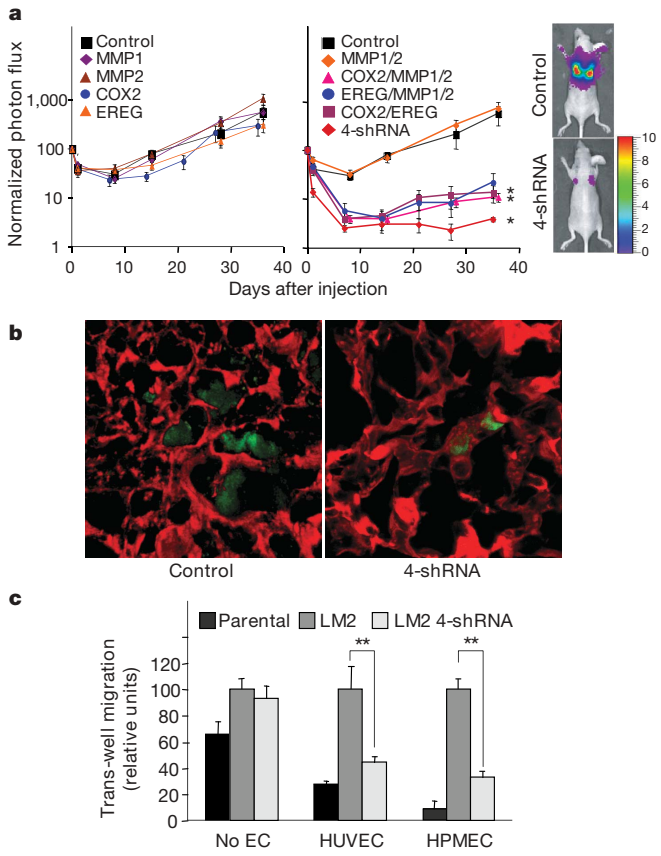


Figure 3 | Genetic inhibition of *EREG*, *MMP1*, *MMP2* and *COX2* prevents metastatic extravasation. **a**, Single and combination knockdown cells were inoculated intravenously into mice. Lung metastasis was measured by bioluminescence and quantified. Left: effects of single knockdowns versus control; right: effects of combination knockdowns versus control with representative bioluminescent images of mice injected with control and 4-shRNA-treated cells at day 35. $n = 5$; error bars indicate s.e.m.; asterisk, $P < 0.05$; based on a two-sided rank-sum test compared to shRNA control LM2 cells. **b**, Control or 4-shRNA-targeted cells were pulsed with cell tracker green and injected into the tail vein of mice. Forty-eight hours after tumour cell inoculation, animals were injected intravenously with rhodamine-conjugated lectin. Whole lungs were then extracted after necropsy and imaged by two-photon confocal microscopy at ×63 magnification. Three-dimensional reconstructed images of tumour cells (green) relative to lung capillaries (red) are shown. **c**, Indicated cell lines were seeded into trans-well inserts with or without (No EC) an endothelial monolayer. Images of cells migrating were captured at ×10 magnification and quantified with ImageJ software. Trans-endothelial migration was performed with either human umbilical vein endothelial cell (HUVEC) or human pulmonary microvascular endothelial cell (HPMEC) monolayers with similar results. $n = 6$–10; error bars indicate s.e.m.; double asterisk, $P < 0.01$; based on a two-sided Student's *t*-test.

Combined drug inhibition of tumour growth and dissemination

These metastasis-promoting activities can also be pharmacologically targeted using previously characterized doses of the anti-EGFR antibody cetuximab[24], the broad-spectrum MMP inhibitor GM6001[25] and the COX2 inhibitor celecoxib[26]. We used an orthotopic model to assess the efficacy of these drug combinations as interventions during the natural formation of lung metastasis from mammary tumours (Fig. 4a). When used as single agents, these drugs minimally inhibited tumour growth of LM2 cells in the mammary glands. However, consistent with our genetic studies, treatment with the cetuximab/celecoxib and cetuximab/celecoxib/GM6001 combinations reduced the rate of primary tumour growth (Fig. 4b). This was accompanied by vascular defects that precipitated tissue hypoxia and ensuing tumour cell apoptosis (Supplementary Fig. 8). This vascular phenotype is also consistent with the ability of cetuximab and celecoxib to inhibit angiogenesis by means of both autocrine and paracrine mechanisms[27,28].

We examined whether the vascular defects elicited by these drugs also resulted in impaired tumour cell intravasation from the primary site. To this end, the presence of circulating tumour cells was assessed by measuring the relative expression of human-specific *GAPDH* in blood from treated mice. Notably, both drug combinations (cetuximab/celecoxib and cetuximab/celecoxib/GM6001) diminished the presence of circulating tumour cells, with the effect of the cetuximab/celecoxib combination reaching statistical significance (Fig. 4c).

Despite the inhibitory effects on intravasation, we noticed that some tumour cells had already colonized the lungs of animals before the initiation of pharmacological treatment (day 24), as seen by immunofluorescent staining (data not shown). Moreover, a significant number of tumour cells were still detectable in the lungs of treated mice (Fig. 4d). Confocal imaging after staining for vascular endothelium and cancer cells showed that vehicle-treated mice harboured large lung metastatic lesions that had efficiently extravasated, whereas mice treated with the cetuximab/celecoxib and cetuximab/celecoxib/GM6001 combinations exhibited a strong bias towards smaller micrometastases that remained trapped within the lung vasculature (Fig. 4e). Digital image quantification of multiple lung sections established statistically significant inhibitory effects for both drug combinations on the overall lung metastatic burden, as well as on the size distribution of the lung metastasis lesions (Fig. 4f, g). GM6001 addition did not increase the inhibitory effects of the cetuximab/celecoxib combination.

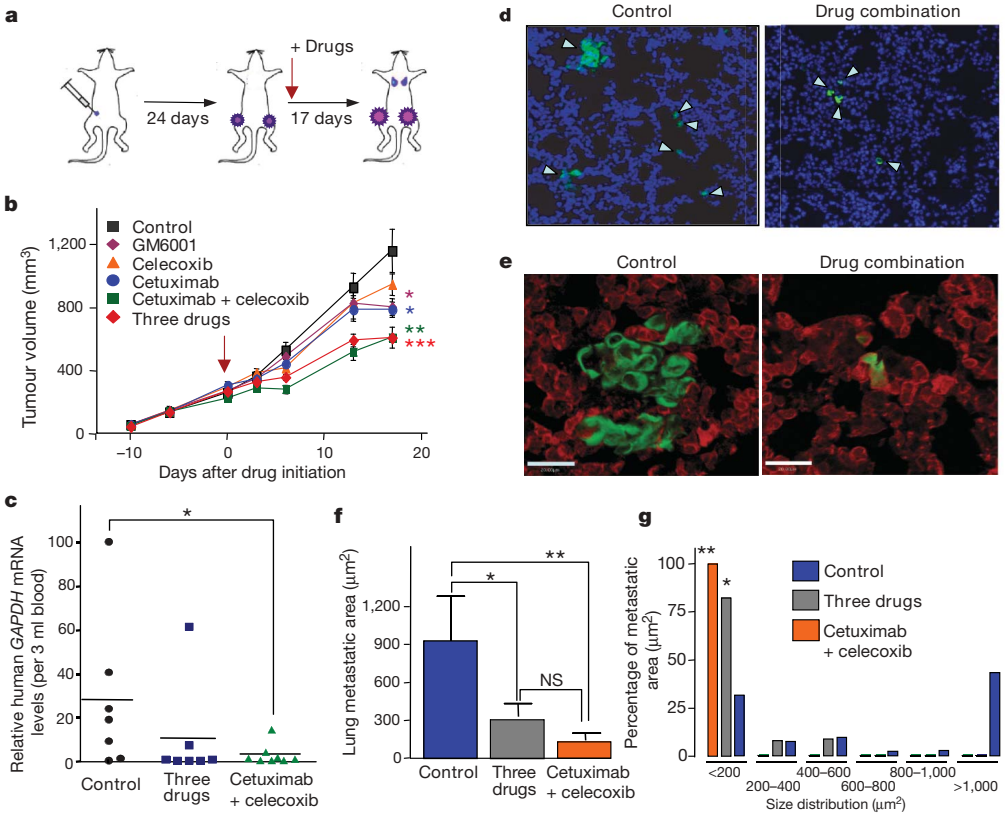


Figure 4 | Pharmacological inhibition of tumour growth and dissemination in orthotopically implanted primary tumours. **a**, Schematic representation of the time course for tumour cell implantation, primary tumour growth and therapeutic treatment to assess effects on primary tumour and lung metastatic progression. Animals were treated with vehicle control, cetuximab, celecoxib, GM6001, or the indicated combinations. **b**, Tumour volume measurements of mice treated with vehicle control or targeted therapies, either individually or in combination. $n = 6$; error bars indicate s.e.m.; asterisk, $P < 0.05$; double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$; based on a two-sided Student's *t*-test, compared to control-treated animals. **c**, Blood from mice was isolated and red blood cells lysed. RNA from the remaining cells was extracted for qRT–PCR. The presence of circulating tumour cells was assessed as a function of human-specific *GAPDH* expression relative to murine *B2m*, in 3 ml of mouse blood perfusate. $n = 7$–8; bars indicate median *GAPDH* expression; asterisk, $P < 0.05$; based on a two-sided Student's *t*-test. **d**, Staining for tumour cells in lung cryosections using a human-specific vimentin antibody. Nuclei were visualized using 4,6-diamidino-2-phenylindole (DAPI). Arrowheads indicate tumour cell clusters in representative ×10 images of lungs from control (left) and animals treated with all three drugs (right). **e**, Confocal imaging at ×63 magnification of tumour cells (vimentin, green) and lung vasculature (lectin, red) from control (left) and treated animals (three drugs, right). Scale bars, 20 μm. **f**, Digital quantification of lung metastatic burden. Fluorescein isothiocyanate (FITC)-stained cancer cells in the lungs were quantified as metastatic lesions, and normalized to the area of nuclear DAPI staining. Shown on the graph is the average lung metastatic area per 5×10^5 μm² DAPI area. $n = 10$; error bars indicate s.e.m. **g**, Percentage of the lung metastatic burden that was occupied by lesions of the indicated sizes. **f**, **g**, Asterisk, $P < 0.05$; double asterisk, $P < 0.01$; NS, not significant, based on a two-tailed Wilcoxon rank-sum test.

Inhibiting lung colonization by metastatic cells

When used as single agents, these drugs also minimally inhibited pulmonary outgrowth of LM2 cells directly inoculated into the tail vein, whereas in combination these drugs prevented metastasis in a manner that mimicked the effects of genetic knockdown (Fig. 5a and Supplementary Fig. 9a). An exception to this trend was the antagonistic interaction between celecoxib and GM6001, perhaps reflecting the fact that the latter is a broad-spectrum MMP inhibitor, likely to affect both pro- and anti-metastatic MMPs[29]. Although the interactions between these agents did not achieve statistical thresholds for synergy, the combination of cetuximab and celecoxib, as well as of cetuximab and GM6001, resulted in supra-additive inhibition of lung metastasis (Supplementary Fig. 9b). Consistent with the genetic inhibition studies, combined pharmacological intervention resulted in the capillary entrapment of the remaining tumour cells even when visualized 4 weeks after tumour cell inoculation (Fig. 5b). Interestingly, the inhibition of extravasation was reversed on discontinuation of drug treatment, with most of the remaining lesions expanding into colonies that invaded the lung parenchyma (Fig. 5b, right panel).

We performed similar experiments with malignant cells freshly obtained from the pleural fluid of two patients with advanced breast cancer and a diagnosis of lung metastasis who were undergoing routine therapeutic procedures at our institution. Carcinoma cells were isolated from these samples based on the epithelial cell marker EpCAM, under institutionally approved protocols[30]. One sample (CN34) was subjected to *in vivo* selection for metastasis-forming cells[12]. As with the MDA-MB-231 cell line, this process yielded a subpopulation (CN34.2A) that expresses high levels of *EREG*, *COX2* and *MMP1* compared with the unsorted CN34 population (Fig. 5c). This was associated with an elevated lung colonizing activity of CN34.2A cells in mice, which could be inhibited by the administration of cetuximab and celecoxib (Fig. 5d, e). A sample from a different patient (CN41) was used without prior experimental selection for metastatic cells. CN41 cells expressed higher levels of *EREG* and *COX2* than did MDA-MB-231 cells (Fig. 5c), and had basal lung colonizing activity that could also be inhibited by the celecoxib/cetuximab combination (Fig. 5d). In all cases the metastasis inhibitory effects of this combination were observed from the earliest stages of metastatic colonization (day 2) and were sustained over 5 weeks (data not shown). These results support the general relevance of these genes in lung colonization, and the ability to interfere with this process by combining their pharmacological inhibitors.

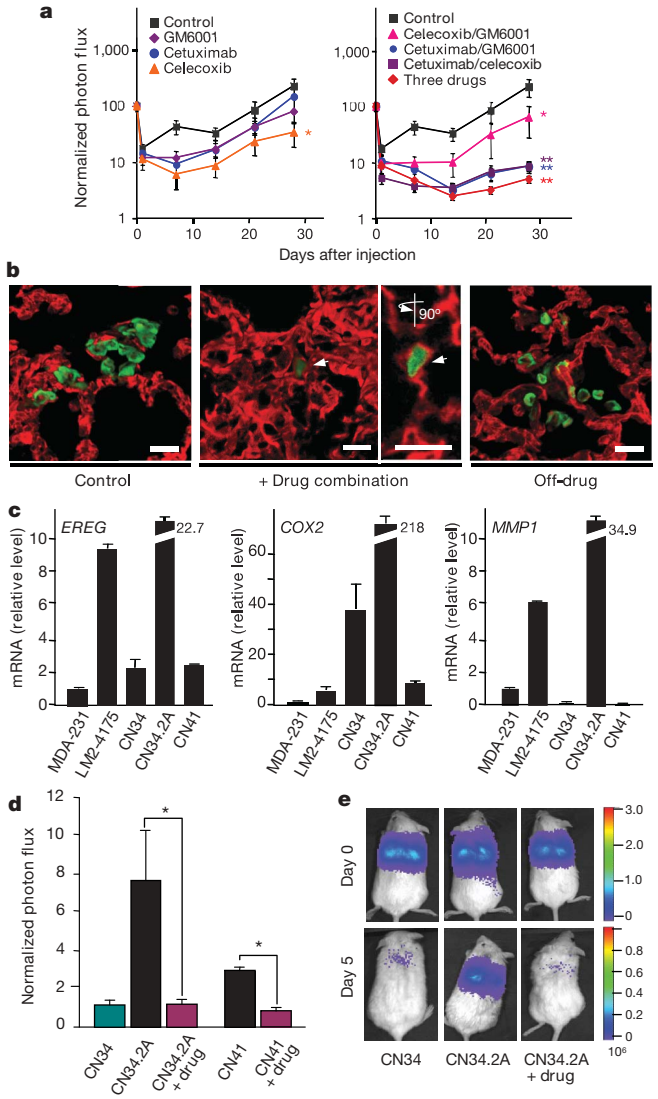


Figure 5 | Targeted inhibition of metastatic extravasation and lung colonization by LM2 and primary malignant cells. **a**, Mice were pre-treated with the indicated agents, 2 days before tumour cell inoculation. After intravenous injections of LM2 cells, drug treatment was maintained and lung metastasis quantified as in Fig. 2. Left: effects of single agents versus control; right: combination drug treatments versus control. $n = 5$; error bars indicate s.e.m.; asterisk, $P < 0.05$; double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$; calculated using a two-sided rank-sum test compared to control treated animals. **b**, Lungs from control or cetuximab/celecoxib/GM6001 (drug combination)-treated mice were collected at day 28. Treatment was terminated in a subset of animals and mice were monitored for an additional 6 weeks. Lungs from these 'off-drug' mice were then collected after necropsy. Lung sections from vehicle-treated, drug-combination-treated and off-drug mice were processed for immunofluorescent detection of tumour-specific vimentin (green) and CD31 (red). Scale bars, 20 μm. **c**, Primary breast carcinoma cell populations CN34 and CN41 were isolated from the pleural effusion of patients. The CN34 derivative, CN34.2A, was obtained by *in vivo* selection of highly metastatic cells in mice. Expression of *EREG*, *COX2* and *MMP1* was assessed by qRT–PCR and compared to parental MDA-MB-231 and LM2 cell lines. $n = 3$; error bars represent 95% confidence interval. **d**, CN34, CN34.2A and CN41 were injected intravenously, mice were treated with cetuximab and celecoxib, and lung colonization was measured via bioluminescence. $n = 5$; error bars indicate s.e.m.; asterisk, $P < 0.05$; based on a two-sided rank-sum test. **e**, Representative luminescence images.

Discussion

Our current findings identify *EREG*, *COX2*, *MMP1* and *MMP2* as a subset of LMS genes that are co-opted by breast cancer cells and reconstitute a multi-functional vascular remodelling programme that promotes metastatic progression. In orthotopic primary tumour assays, these factors collectively mediate pathological angiogenic progression, with an ensuing increase in vascular permeability and tumour cell intravasation. Remarkably, this set of genes is also required to breach the lung vasculature and enable extravasation of cancer cells on dissemination of these cells from mammary tumours to the lungs. Although the individual targeting of these mediators was insufficient to prevent such biological activities, their combined inhibition resulted in profound reductions in sequential steps of metastatic progression.

The engagement of a common set of factors in distinct steps of metastasis qualifies them as 'metastasis progression' genes, which we define as genes that fulfil certain rate-limiting functions in primary tumour growth and other specific functions in metastatic colonization. Metastasis progression genes are thus distinguished from oncogenes that fulfil cell-autonomous transforming functions through the course of malignant disease, and from 'metastasis virulence' genes, defined as those genes that participate in metastatic colonization but not in primary tumour development.

Despite the major clinical advances provided by cytotoxic, hormonal and targeted therapies, the median survival after diagnosis of metastatic breast cancer with visceral organ involvement remains less than 2 years[31–33]. For the most part, the target of commonly used chemotherapy drugs continues to be the proliferation of cancer cells. As the molecular understanding of the biological functions necessary for metastasis increases, it may become possible to develop anti-metastatic strategies that target not only the intrinsic growth of disseminated tumour cells, but also their necessary interactions with newly adopted microenvironments. Our current observations demonstrate that inhibition of EGFR and COX2 can abate lung metastatic progression in a clinically relevant model of breast cancer. Collectively, these results identify extravasation as an essential step in metastatic progression that can be inhibited by combinatorial therapies formulated on the basis of biological insights.

METHODS SUMMARY

MDA-MB-231 and its lung metastatic derivative LM2-4175 have been described previously[12,17]. CN34 and CN41 carcinoma cells were isolated from the pleural effusion of patients with metastatic breast cancer treated at our institution on obtained written consent in accordance with IRB regulations, as previously described[30]. Metastatic cell subpopulations were obtained by *in vivo* selection in mice[12]. Knockdown of *MMP1*, *MMP2* and *COX2* was achieved using pRetroSuper technology[12]. For *EREG* targeting, an alternative vector was used (pSM2 derivative), which expresses the short hairpin embedded in a larger microRNA sequence[34]. All animal work was done in accordance with the MSKCC Institutional Animal Care and Use Committee. BALB/c nude and NOD/SCID female mice (NCI) age-matched between 5–7 weeks were used for xenografting studies. For inhibitor studies, 1 mg cetuximab was injected intraperitoneally biweekly, which yields plasma drug concentrations within the corresponding range in cetuximab-treated cancer patients[35]. GM6001 (Ryss Lab) was injected intraperitoneally at a dose of 2 mg kg^{-1} daily, which is efficacious in preclinical mouse models[25]. Celecoxib (LKT laboratories) was mixed with a powdered rodent chow diet (Research Diets) at a concentration of 1,000 parts per million, and provided continuously during the course of the experiment. Celecoxib serum concentrations in mice treated within this range are clinically attainable and sufficient to inhibit inflammation and prostaglandin synthesis in humans[22,26,36]. The Methods section provides additional information including cell culture, malignant cell isolation from pleural fluids, generation of retroviral gene-knockdown vectors, infections and transfections, analysis of RNA and protein expression, trans-endothelial migration assays, animal inoculation and bioluminescence imaging, sources and use of pharmacological inhibitors, *in vivo* intravasation assays, tumour histological and immunohistochemical analyses, vascular permeability assays, extravasation visualization, and image quantification.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 December 2006; accepted 21 March 2007.

1. Fidler, I. J. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nature Rev. Cancer* **3**, 453–458 (2003).

2. Hynes, R. O. Metastatic potential: generic predisposition of the primary tumor or rare, metastatic variants—or both? *Cell* **113,** 821–823 (2003).
3. Gupta, G. P. & Massagué, J. Cancer metastasis: building a framework. *Cell* **127,** 679–695 (2006).
4. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406,** 747–752 (2000).
5. van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347,** 1999–2009 (2002).
6. Chang, H. Y. *et al.* Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol.* **2,** E7 (2004).
7. Paik, S. *et al.* A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351,** 2817–2826 (2004).
8. Wang, Y. *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365,** 671–679 (2005).
9. Bild, A. H. *et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439,** 353–357 (2006).
10. Massagué, J. Sorting out breast-cancer gene signatures. *N. Engl. J. Med.* **356,** 294–297 (2007).
11. Bernards, R. & Weinberg, R. A. A progression puzzle. *Nature* **418,** 823 (2002).
12. Minn, A. J. *et al.* Genes that mediate breast cancer metastasis to lung. *Nature* **436,** 518–524 (2005).
13. Minn, A. J. *et al.* Lung metastasis genes couple breast tumor size and metastatic spread. *Proc. Natl Acad. Sci. USA* **104,** 6740–6745 (2007).
14. Cailleau, R., Young, R., Olive, M. & Reeves, W. J. Jr. Breast tumor cell lines from pleural effusions. *J. Natl Cancer Inst.* **53,** 661–674 (1974).
15. Gupta, G. P. *et al.* Identifying site-specific metastasis genes and functions. *Cold Spring Harb. Symp. Quant. Biol.* **70,** 1–10 (2005).
16. Kang, Y. *et al.* A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* **3,** 537–549 (2003).
17. Minn, A. J. *et al.* Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors. *J. Clin. Invest.* **115,** 44–55 (2005).
18. Wary, K. K., Thakker, G. D., Humtsoe, J. O. & Yang, J. Analysis of VEGF-responsive genes involved in the activation of endothelial cells. *Mol. Cancer* **2,** 25 (2003).
19. Carmeliet, P. & Jain, R. K. Angiogenesis in cancer and other diseases. *Nature* **407,** 249–257 (2000).
20. Krysan, K. *et al.* Prostaglandin E2 activates mitogen-activated protein kinase/Erk pathway signaling and cell proliferation in non-small cell lung cancer cells in an epidermal growth factor receptor-independent manner. *Cancer Res.* **65,** 6275–6281 (2005).
21. Dannenberg, A. J., Lippman, S. M., Mann, J. R., Subbaramaiah, K. & DuBois, R. N. Cyclooxygenase-2 and epidermal growth factor receptor: pharmacologic targets for chemoprevention. *J. Clin. Oncol.* **23,** 254–266 (2005).
22. Howe, L. R. *et al.* HER2/neu-induced mammary tumorigenesis and angiogenesis are reduced in cyclooxygenase-2 knockout mice. *Cancer Res.* **65,** 10113–10119 (2005).
23. Pai, R. *et al.* Prostaglandin E2 transactivates EGF receptor: a novel mechanism for promoting colon cancer growth and gastrointestinal hypertrophy. *Nature Med.* **8,** 289–293 (2002).
24. Goldstein, N. I., Prewett, M., Zuklys, K., Rockwell, P. & Mendelsohn, J. Biological efficacy of a chimeric antibody to the epidermal growth factor receptor in a human tumor xenograft model. *Clin. Cancer Res.* **1,** 1311–1318 (1995).
25. Gijbels, K., Galardy, R. E. & Steinman, L. Reversal of experimental autoimmune encephalomyelitis with a hydroxamate inhibitor of matrix metalloproteases. *J. Clin. Invest.* **94,** 2177–2182 (1994).
26. Jacoby, R. F., Seibert, K., Cole, C. E., Kelloff, G. & Lubet, R. A. The cyclooxygenase-2 inhibitor celecoxib is a potent preventive and therapeutic agent in the min mouse model of adenomatous polyposis. *Cancer Res.* **60,** 5040–5044 (2000).
27. Gately, S. & Li, W. W. Multiple roles of COX-2 in tumor angiogenesis: a target for antiangiogenic therapy. *Semin. Oncol.* **31,** 2–11 (2004).
28. Karashima, T. *et al.* Inhibition of angiogenesis by the antiepidermal growth factor receptor antibody ImClone C225 in androgen-independent prostate cancer growing orthotopically in nude mice. *Clin. Cancer Res.* **8,** 1253–1264 (2002).
29. Overall, C. M. & Kleifeld, O. Tumour microenvironment—opinion: validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy. *Nature Rev. Cancer* **6,** 227–239 (2006).
30. Gomis, R. R., Alarcon, C., Nadal, C., Van Poznak, C. & Massagué, J. C/EBPβ at the core of the TGFβ cytostatic response and its evasion in metastatic breast cancer cells. *Cancer Cell* **10,** 203–214 (2006).
31. Giordano, S. H. *et al.* Is breast cancer survival improving? *Cancer* **100,** 44–52 (2004).
32. Solomayer, E. F., Diel, I. J., Meyberg, G. C., Gollan, C. & Bastert, G. Metastatic breast cancer: clinical course, prognosis and therapy related to the first site of metastasis. *Breast Cancer Res. Treat.* **59,** 271–278 (2000).
33. Baselga, J. & Norton, L. Focus on breast cancer. *Cancer Cell* **1,** 319–322 (2002).
34. Silva, J. M. *et al.* Second-generation shRNA libraries covering the mouse and human genomes. *Nature Genet.* **37,** 1281–1288 (2005).
35. Luo, F. R. *et al.* Correlation of pharmacokinetics with the antitumor activity of Cetuximab in nude mice bearing the GEO human colon carcinoma xenograft. *Cancer Chemother. Pharmacol.* **56,** 455–464 (2005).
36. Niederberger, E. *et al.* Celecoxib loses its anti-inflammatory efficacy at high doses through activation of NF-κB. *FASEB J.* **15,** 1622–1624 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Minn, D. Padua, C. Van Poznak, L. Norton, C. Hudis, Y. Pylayeva, P. Gupta, T. Westbrook and Z. Lazar for insightful discussions and technical suggestions. We also thank S. Tulley and members of the Molecular Cytology Core Facility for expert technical assistance. J.M. was funded by a National Institutes of Health grant, and by grants of the W. M. Keck Foundation and the Kleberg Foundation. G.P.G. is supported by an NIH Medical Scientist Training Program grant and a Department of Defense Breast Cancer Research Program pre-doctoral award. D.X.N. is a Berlex postdoctoral fellow of the Damon Runyon Cancer Research Foundation. A.C.C. was supported by an ASCO Young Investigator Award and by the Charles A. Dana Foundation. J.M. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions J.M. designed and supervised experiments. G.P.G., D.X.N. and A.C.C. designed experiments; G.P.G., D.X.N., A.C.C., P.D.B. and J.Y.K. performed experiments; C.N. and R.R.G. isolated metastatic cells from clinical samples; K.T.-M. supervised histological and confocal microscopy imaging; J.M., G.P.G., D.X.N. and A.C.C. analysed data and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.M. (j-massague@ski.mskcc.org).

METHODS

Cell culture. MDA-MB-231 and its lung metastatic derivative LM2-4175 have been described previously[12,17]. All tumour cell lines were cultured in DMEM supplemented with 10% FBS, glutamine, penicillin, streptomycin and fungizone. CN34 and CN41 carcinoma cells were isolated from the pleural effusion of patients with metastatic breast cancer treated at our institution upon written consent obtained following IRB regulations as previously described[30]. Briefly, pleural effusion samples were centrifuged at 1,000 r.p.m. for 10 min, cell pellets were re-suspended in PBS and treated with ACK lysis buffer to lyse blood cells. A fraction of these cells underwent negative selection to remove leukocytes ($CD45^+$ and $CD15^+$ cells), and EpCam-positive cells were sorted from the population upon recovery in tissue culture for 24 h. HUVEC (ScienCell) endothelial cells were cultured in complete ECM media (ScienCell), whereas primary human pulmonary microvascular endothelial cells (HPMECs, Cambrex) were cultured in EGM-2 (Cambrex). HUVECs and HPMECs were used between passages 3–6. The retroviral packaging cell line GPG29 was maintained in DMEM containing 10% FBS supplemented with puromycin, G418, doxycycline, penicillin, streptomycin and fungizone. Transfections were done using standard protocols with Lipofectamine 2000 (Invitrogen). After transfection, GPG29 cells were cultured in DMEM containing 10% FBS and sodium pyruvate.

Generation of retrovirus and knockdown cells. Knockdown of *MMP1*, *MMP2* and *COX2* was achieved using pRetroSuper technology targeting the following 19-nucleotide sequences: 5′-AGCGGAGAAATAGTGGCCC-3′ (*MMP1*), 5′-GGACGGACTCCTGGCTCAT-3′ (*MMP2*) and 5′-GGGCTGTCCCTTTACTTCA-3′ (*COX2*). Knockdown of *IL13RA2*, *SPARC* and *VCAM1* was achieved as previously described[12]. For *EREG* targeting, an alternative vector was used (pSM2 derivative), which expresses the short hairpin embedded in a larger microRNA sequence[34]. The two target sequences used in the *EREG* gene were 5′-CCCAATATATTCTGACCGTTAA-3′ and 5′-ACCACAAATGCATAAATGCATA-3′. To produce retrovirus for combination knockdown, multiple hairpin vectors were transfected as pools into the GPG29 amphotropic packaging cell line. Viruses were collected 48 and 72 h after transfection, filtered, and concentrated by ultracentrifugation. Concentrated retrovirus was used to infect cells in the presence of 8 μg ml^{-1} polybrene, typically resulting in a transduction rate of over 80%, and infected cells were selected with puromycin. Because the total amount of plasmid DNA used for co-transfection of multiple hairpins was the same as that used for single hairpin transfection, the combination knockdown retroviral titres were similar to titres attained when generating single knockdown virus. Moreover, we have demonstrated that up to four different vectors can be delivered and expressed efficiently in MDA-MB-231 cells using this protocol. Thus, knockdown cells obtained after drug selection were injected as a pooled population without subcloning. To generate knockdown-rescue cell lines, we used a similar method to produce virus encoding complementary DNAs for overexpression of the RNAi-targeted genes, along with a hygromycin selectable marker. Combination overexpressing retrovirus was used to super-infect previously generated knockdown cells that were subsequently selected with hygromycin.

Analysis of mRNA and protein expression. Total RNA from subconfluent MDA-MB-231 cells was collected and purified using the RNeasy kit (Qiagen). Five-hundred nanograms of total purified RNA was subjected to a reverse transcriptase reaction according to the SuperScript III first-strand synthesis system (Invitrogen). cDNA corresponding to approximately 5 ng of starting RNA was used for each of four replicates for quantitative PCR. Human *EREG*, *MMP1*, *MMP2*, *COX2* and *β2-microglobulin* (as an endogenous control) were amplified with commercially designed Taqman gene expression assays (Applied Biosystems) and the Taqman universal PCR master mix (Applied Biosystems). Quantitative expression data were acquired and analysed using an ABI Prism 7900HT Sequence Detection System (Applied Biosystems). For immunoblotting, cells were washed with PBS and lysed in RIPA buffer (50 mM Tris-HCl, pH 7.4, 1% NP-40, 0.25% Na-deoxycholate, 150 mM NaCl, 1 mM EDTA) supplemented with 50 mM NaF, 20 mM β-glycerophosphate, and complete protease inhibitor cocktail (Roche). Proteins were separated by SDS–PAGE and transferred to nitrocellulose membranes (Bio-Rad) that were immunoblotted with mouse monoclonal antibodies that recognize COX2 (Cayman Chemicals) and α-tubulin (Sigma). To facilitate detection of endogenous COX2 protein, cells were also pre-treated with 20 ng ml^{-1} TNF-α for 5 h before lysing (R&D Systems). For analysis of secreted protein expression, cells were plated in triplicate at 90% confluency in 12-well plates, incubated in DMEM 0.2% FBS, and conditioned media was collected 72 h later. Media was cleared of cells by centrifuging at 2,000 r.p.m. for 5 min. Pro-MMP1, pro-MMP2 and VEGF-165 concentrations were analysed in conditioned media using ELISA kits (R&D Systems).

Animal studies. All animal work was done in accordance with a protocol approved by the MSKCC Institutional Animal Care and Use Committee. BALB/c nude and NOD/SCID female mice (NCI) age-matched between 5–7 weeks were used for xenografting studies. For primary tumour analysis, 1×10^6 viable single cells were re-suspended in a 1:1 mixture of PBS and growth-factor-reduced Matrigel (BD Biosciences) and injected orthotopically into mammary gland four in a total volume of 50 μl as previously described[12]. Primary tumour growth rates were analysed by measuring tumour length (L) and width (W), and calculating tumour volume based on the formula $\pi LW^2/6$. For experimental metastasis assays, 2×10^5 cells were re-suspended in 0.1 ml PBS and injected into the lateral tail vein. Lung metastatic progression was monitored and quantified using non-invasive bioluminescence as previously described[12].

Pharmacological inhibitors. Cetuximab (ImClone) was obtained from the MSKCC pharmacy. For inhibitor studies, 1 mg cetuximab was injected intraperitoneally bi-weekly. Injection with cetuximab at doses between 0.25 mg and 1 mg per injection achieve plasma drug concentrations within the corresponding range in cetuximab-treated cancer patients[35]. GM6001 (Ryss Lab) was injected intraperitoneally at a dose of 2 mg kg^{-1} daily, which has previously been shown to be efficacious in preclinical mouse models[25]. Celecoxib (LKT laboratories) was mixed with a powdered rodent chow diet (Research diets) at a concentration of 1,000 parts per million (1 g celecoxib per 1 kg chow), and provided continuously during the course of the experiment. Previous pharmacokinetic studies demonstrate that celecoxib serum concentration in mice treated within this range are clinically attainable and sufficient to inhibit inflammation and prostaglandin synthesis in humans[22,26,36].

Trans-endothelial migration. HUVECs or primary HPMECs were seeded into collagen-coated trans-well inserts (1 μm pore size, BD Falcon) at 100,000 cells per well, and allowed to grow to confluence for 4 days. Tumour cells were pulsed with 5 μM cell tracker green (Invitrogen) for 30 min before being conditioned overnight in 0.2% FBS ECM media without growth factors. The next day, 50,000 tumour cells were seeded into trans-well inserts with or without a confluent endothelial monolayer, and the wells were fixed in 4% paraformaldehyde after 10 h. Cells on the apical side of each insert were scraped off and the trans-well membrane mounted onto slides. Migration to the basolateral side of the membrane was visualized with a Zeiss Axioplan2 immunofluorescent microscope at ×10 magnification. Pictures of 6–10 random fields across three replicate wells were captured for quantification using ImageJ software (NIH). In general, 150–200 counts per field of LM2 cells were seen to migrate in the absence of a monolayer, whereas 50 counts per field were seen to migrate through an endothelial barrier. Migration of the indicated lines was plotted as a percentage of migrating LM2 control cells.

Intravasation. Drug-treated mice were perfused with 5 ml PBS through the left ventricle. Three millilitres of blood perfusate was collected from the atrium and lysed two times using ACK lysis buffer (Cambrex). Total RNA was extracted from the remaining cells and used for qRT–PCR as described above. The presence of human circulating tumour cells was determined by the relative expression of human *GAPDH* normalized to murine β2-microglobulin.

Tumour and lung immunostaining. Mice were killed and perfused with PBS and 4% paraformaldehyde through the left ventricle, before tumours were extracted, fixed and paraffin-embedded. Immunohistochemical staining for CD31 (Santa Cruz), phospho-histone H3 (Upstate) and cleaved caspase-3 (Cell Signaling) was performed on paraffin-embedded tumour sections by the MSKCC Molecular Cytology Core Facility. Brightfield microscopic images were collected using an Axioplan2 microscopy system (Zeiss). Tumour cell proliferation (pH3) and apoptosis (cleaved caspase-3) were quantified using a combination of Adobe Photoshop (Adobe) and ImageJ software (NIH). In brief, the colour-picker function was used to identify manually the most darkly stained region of interest, with a constant fuzziness factor. The selected regions were feathered and expanded in a uniform manner, and thresholded into binary images, which were subsequently analysed in ImageJ. Morphometric analysis of CD31-stained vessels was achieved with Photoshop and Image Processing Tool Kit (Reindeer Graphics Inc.) based on a previously described protocol[37]. Angiogenic properties were then scored as a function of vessel density, average vessel length, average number of branch points per vessel, and lumen formation. On average, immunohistochemistry quantification was performed by taking pictures from five random fields per tumour, imaging at least three tumours per sample set. For immunofluorescence, tumours were fixed and frozen in OCT. Pericyte coverage of vessels was identified by double staining for the pericyte marker NG2 (Chemicon) and CD31 endothelial cell marker (BD Biosciences Pharmingen), followed by detection with fluorescently conjugated secondary antibodies (Jackson Immunoresearch). Permeability of tumour blood vessels was assessed by intravenous injection of rhodamine-conjugated dextran (70 kDa, Invitrogen) at 2 mg per 20 g body weight. After 1 h, mice were killed, tumours extracted, and 30-μm sections examined by fluorescence microscopy for vascular leakage. Immunofluorescent staining for pimonidazole adducts in

primary tumours was performed according to the Hypoxyprobe-1 staining kit (Chemicon).

To observe metastatic extravasation within the first 48 h of circulatory entry, tumour cells were labelled with 5 μM of cell tracker green (Invitrogen) for 1 h and inoculated into mice. Before sacrifice, mice were injected intravenously with rhodamine-conjugated lectin (Vector Laboratories) to stain the lung vasculature. Lungs were perfused with PBS, inflated through intra-tracheal injection, and extracted *en bloc*. Whole lungs were then scanned by two-photon confocal microscopy at ×63 using a Leica TCS SP2 microscope (DM IRE2 inverted stand). Representative three-dimensional images of stained capillaries and tumour cells were processed using Volocity v.3.6 (Improvision). To examine extravasating cells in the drug-treated mice, an alternative protocol was used. In this case, tumour cells did not retain cell tracker label during the extended time frame of the experiment and were alternatively co-stained with a monoclonal antibody that selectively detects human vimentin (Novocastra); anti-CD31 antibody (BD Biosciencies Pharmingen) was used to visualize lung capillaries. After fluorescent secondary antibody incubation, images were captured with a Leica TCS SP2 microscope (DMRXA2 upright stand) and processed using Volocity (Improvision). Quantification of lung metastasis was performed by creating montage images of whole-lung sections at ×10 magnification using an Axiovert 200M imaging system equipped with a motorized inverted stand (Zeiss). Image quantification was performed as described above using a combination of Adobe Photoshop (Adobe) and ImageJ software (NIH).

37. Wild, R., Ramakrishnan, S., Sedgewick, J. & Griffioen, A. W. Quantitative assessment of angiogenesis and tumor vessel architecture by computer-assisted digital image analysis: effects of VEGF-toxin conjugate on tumor microvessel density. *Microvasc. Res.* **59**, 368–376 (2000).

LETTERS

A laboratory demonstration of the capability to image an Earth-like extrasolar planet

John T. Trauger[1] & Wesley A. Traub[1]

The detection and characterization of an Earth-like planet orbiting a nearby star requires a telescope with an extraordinarily large contrast at small angular separations. At visible wavelengths, an Earth-like planet would be 1×10^{-10} times fainter than the star at angular separations of typically 0.1 arcsecond or less[1,2]. There are several proposed space telescope systems that could, in principle, achieve this[3–6]. Here we report a laboratory experiment that reaches these limits. We have suppressed the diffracted and scattered light near a star-like source to a level of 6×10^{-10} times the peak intensity in individual coronagraph images. In a series of such images, together with simple image processing, we have effectively reduced this to a residual noise level of about 0.1×10^{-10}. This demonstrates that a coronagraphic telescope in space could detect and spectroscopically characterize nearby exoplanetary systems, with the sensitivity to image an 'Earth-twin' orbiting a nearby star.

Coronagraphs are not new to astronomy[7], but only recently has the concept been extended to the imaging of Earth-like extrasolar planets (exoplanets) from space[8]. In space, free of the blurring effects of atmospheric turbulence, a coronagraph must further suppress light diffracted from the edges of the primary mirror (the well-known Airy rings[9]) as well as a surrounding field of speckles due to irregularities in the surface figure of the optics. If left uncontrolled, these effects spread starlight across the focal plane, completely overwhelming any faint planet image. The scattered light in a high-quality optical telescope in space (for example, the Hubble Space Telescope, HST) at a typically close angular separation of $4\lambda/D$ is about 10^{-3} times the peak intensity of the central star (here λ is wavelength, and D is the diameter of the telescope primary mirror). This angular separation roughly corresponds to an Earth-twin, at maximum angular separation, orbiting a star 10 pc distant, as observed by a 4-m telescope at 500 nm wavelength. However, the scattered background is about 10^7 times greater than an Earth-twin; it must be suppressed to obtain a clean exoplanet signal.

In practice, diffracted light from the edges of the primary mirror can be removed by a variety of well-studied coronagraph configurations[10–16], each with its own specific characteristics and limitations, including efficiency, spectral bandwidth and complexity. The coronagraph can be combined with a telescope in a single spacecraft[3] to form a system that can be tested for end-to-end performance before launch. This is a significant advantage in cost, complexity and risk when compared to other actively studied mission concepts[4–6], which all require multiple spacecrafts orbiting in precisely constrained formations—a technology that cannot be validated on the ground and has yet to be proven in space. Here we demonstrate suppression of diffracted light with a Lyot-type coronagraph, and suppression of the speckle field with a wavefront sensing and control technique that is applicable to all coronagraph types.

Our laboratory apparatus, the High Contrast Imaging Testbed[17] (HCIT), is detailed in Supplementary Information, where we describe the optical elements, the concept of a band-limited Lyot coronagraph, the use of a precision deformable mirror (DM) to create a high-contrast dark field of view, a demonstration of the speckle nulling procedure that senses and corrects errors in the optical wavefront, and a discussion of spectral bandwidth. Enclosed in a space-like environment inside a vibration-isolated and thermally controlled vacuum chamber at room temperature, this system captures the essential optical features of a space coronagraph.

We report here the results of two experiments with the HCIT, namely, a 'snapshot' and a 'movie'. The snapshot experiment simulates a single exposure of a star and exoplanet system by a space coronagraph, allowing a view of a one-sided region near the star. The movie experiment simulates a series of snapshots taken as the space coronagraph is rotated about the line of sight to the star, thereby allowing a search for exoplanets in an annular region around the star.

In the snapshot experiment, the image of the star is centred on the coronagraph mask and the DM is commanded so as to minimize the speckle intensity in the target field of view. This experiment used a simulated star, but no simulated planet. By offsetting the star to a clear part of the mask, we record what a planet would look like (Fig. 1a). With the star centred, a dark target field appeared (Fig. 1b). If a simulated planet had been present in this D-shaped field, it would have appeared as a bright spot resembling Fig. 1a. Quantitatively, the data plotted in Fig. 1c show: curve A, the azimuth-averaged intensity of the star in the target field, without the focal plane mask present; curve B, the azimuth-averaged intensity of the star in the target field, with the focal plane mask, and with the DM set to minimize the average intensity in this field; and curve C the same as B except before DM correction. As shown in curve B, the average intensity is about 6×10^{-10} times the peak intensity of the star in a field between $4\lambda/D$ and $10\lambda/D$ on one side of the star. This snapshot experiment shows that the present apparatus, in a single exposure, is capable of suppressing both the diffracted and scattered light around a star, down to a contrast level that is slightly better than a Jupiter, but not quite as low as an Earth[18].

This experiment validates many, but not all, of the critical elements of an actively corrected space coronagraph for exoplanet imaging. It is a simple, stable coronagraph configuration, operating in a space-like environment, with contrast performance that can be accurately modelled end-to-end using the known characteristics of the optical elements. It illustrates a robust method of optical wavefront sensing and control that requires only the DM and a sensor in the focal plane to analyse the image of a star. It shows that current DMs are capable of suppressing scattered light to contrast levels and separations representative of a planet-finding mission and, as the movie experiment shows below, that the precision DM settings remain stable over periods of hours or more without feedback.

However, the experiment lacks a simulator for a large telescope structure in space. We briefly consider the flight system characteristics

[1]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109, USA.

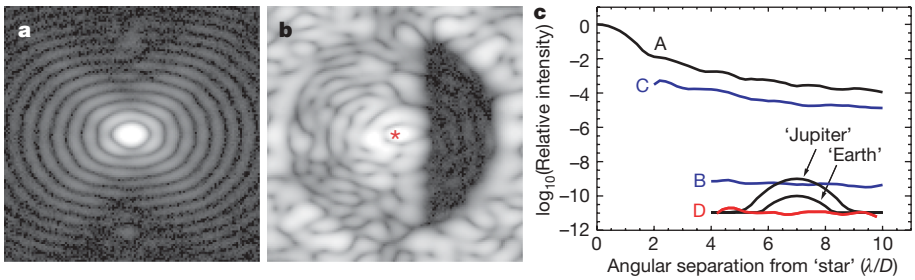


Figure 1 | Representative coronagraph images and intensity profiles. **a**, The appearance of a 'planet' offset from the star (and therefore not occulted by the coronagraph mask). The horizontal elongation of the diffraction rings is a result of the Lyot aperture (see Supplementary Information). **b**, The high-contrast coronagraph 'dark half-field' to the right of the masked central star (at the location of the red asterisk). **c**, Plots of relative intensity versus angular separation from the 'star'. Curve A, data from **a**; curve B, data from **b**; curve C, the coronagraph contrast before wavefront correction (with the DM nominally flat); curve D, the r.m.s. background obtained by roll deconvolution of a set of coronagraph images (obtained by subtracting the azimuthally averaged background in Fig. 2c). Intensity profiles for a nominal Earth and Jupiter in reflected starlight are included for reference. These are equal in intensity to the planets in Fig. 2, but centred at a separation of $7\lambda/D$ for clarity, and standing above the roll-deconvolved r.m.s. background. Images and intensity curves are displayed on logarithmic scales.

needed to carry this coronagraph concept forward to a viable space mission. System considerations will include our ability to manufacture an optical system to the required tolerances; our ability to make the necessary alignments and adjustments in orbit with demonstrated wavefront sensing and control methods; and our ability to maintain, within acceptable limits, critical alignments that may vary on timescales shorter than the wavefront sensing and control cycle (for example, less than 90 min). In Table 1, we list representative engineering tolerances for the most important sources of wavefront error in a 4-m telescope with a Lyot coronagraph.

The experiment was performed in polarized, narrowband laser light rather than unpolarized, continuum starlight filtered to a 10–20% ($\delta\lambda/\lambda$) bandwidth as would be required for photometric studies in astronomy. As we discuss further in Supplementary Information, our coronagraph is insensitive to polarization, working equally well in polarized and unpolarized light. We are currently addressing the question of spectral bandwidth in stages. Early experiments with 2% bandwidth yield a contrast of 1×10^{-9} using the same coronagraph masks and speckle nulling procedure as above (B. Kern, personal communication), and models indicate this can be improved to the contrast levels seen in narrowband light. Initial experiments with 10% bandwidth light, again with the same coronagraph mask and speckle nulling procedure, produced a contrast of 4×10^{-9}. Our model predicts that contrast with 10% bandwidth will be improved by about an order of magnitude using new coronagraph masks now being manufactured with standard techniques and common materials. This is an active area of development and a pathfinder for the ultimate space mission, as described in greater detail in Supplementary Information.

In the movie experiment, to simulate a coronagraph operating in space, we continuously repeated the snapshot experiment 480 times over a period of 5 h. The apparatus was very stable during this period; it was not adjusted in any way between exposures. The background speckle field evolved slowly, owing to room temperature changes and mechanical relaxation, much as might be expected in a coronagraph in space.

The movie experiment demonstrates the process of planet discovery with a space coronagraph, as follows. As shown in Fig. 2a, we hypothesize a space coronagraph that is aimed at a nearby star (centred behind the asterisk), with the background starlight suppressed in the target field, as in the snapshot experiment. We assume that the space coronagraph is rotated in small increments, so that the dark target field successively covers regions that ultimately fill the complete annular region between angular separations of $4\lambda/D$ and $10\lambda/D$ on the sky. For illustration, we show 12 discrete 30° steps in Fig. 2a, but in the computation we use 48 steps of 7.5° each, and in each step we co-add 10 sequential exposures, using a total of 480 exposures.

Lacking a simulated planet, we added attenuated copies of the star to each exposure, at the sky locations shown in Fig. 2a by the three coloured objects. This procedure is valid because we have previously shown that the presence of a planet in the speckle field of a star has no effect whatsoever on our wavefront correction algorithm[19]. The orbital positions were chosen artificially so that the projected planetary system could be captured in a single image here.

Table 1 | Critical engineering tolerances for a flight system

Significant sources of wavefront error	r.m.s. tolerance	Timescale	HST comparison
Precision of the DM settings*	0.014 nm*	>90 min	NA (no DM)
Telescope pointing jitter			
Body pointing of telescope (spacecraft momentum wheels)	35 mas	Active (0.1 Hz)	3–5 mas
Star on the coronagraph mask (fine steering mirror)	2.5 mas	Active (10 Hz)	NA (no FSM)
Surface of the primary mirror, by spatial scale across the PM			
$D/4$ and larger (specification, stability)	8 nm, 1.6 nm	Continuous	9 nm, unknown
$D/50$ to $D/4$ (polish and quilting artefacts, stability)	7 nm, 0.014 nm	>90 min	7 nm, similar
Particulate contamination (dust and micro-meteor pits)	Level 750	Years	Level 750 or better
Coating reflectance uniformity ($D/50$ to $D/4$)	0.25%	Years	Unknown
Surface quality of small optics (specification, stability)	1 nm, 0.014 nm*	>90 min	8 nm, similar
Telescope alignment (separation of the SM and PM)	1,000 nm	Continuous	2,000 nm

Representative engineering tolerances are shown for the dominant sources of wavefront error in a 4-m exoplanet imaging telescope with a Lyot coronagraph, and an HST comparison. The system architecture balances difficult requirements across the flight system, based on commercially available components and trades between competing engineering parameters, and a relaxation of optical tolerances using the DM for active wavefront correction. The tabulated numbers were derived by scaling from the (1.5-m) Eclipse[23] and TPF-C (8-m) SpeckleCam[22] engineering point designs. For the purpose of illustration, engineering tolerances are expressed in terms of their effect, δC, on the contrast metric C for the system overall. The cumulative effect of various independent sources of wavefront error can be approximated as the linear sum of their individual δC allocations, a computation that is equivalent to a quadratic sum of random, uncorrelated wavefront errors. With this understanding, each tolerance in the table has been scaled to contribute a δC no greater than about 1×10^{-11} (roughly one-tenth the brightness of an exo-Earth) to the overall contrast error budget. Some of these drift slowly over time and may be regarded as static errors, to be swept away with each new cycle of speckle nulling. Others, varying on shorter timescales, must be actively maintained within acceptable limits during the time between nulling cycles. In practice, a particular tolerance could be allocated a larger fraction of the overall contrast budget than shown here, but only at the expense of tighter requirements elsewhere in the system. Tolerances are presented along with expected timescales for significant variations (see Supplementary Information for further explanation of these tabulated values). NA, not applicable; FSM, fine steering mirror; PM, SM, telescope primary and secondary mirrors.
* Stability tolerances demonstrated in this Letter.

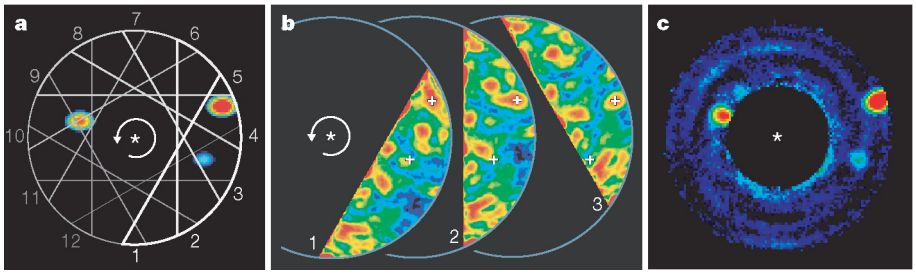


Figure 2 | Laboratory images demonstrate contrast at levels required to detect an Earth-twin. a, Three planet images are shown on the sky. The planets are copies of the measured star but reduced in intensity by factors of (10, 5 and 1) $\times 10^{-10}$, corresponding to the typical intensities of Jupiter, half-Jupiter and Earth, respectively. The Earth-twin is at about 4 o'clock, and the Jupiter-twin at 2 o'clock. The D-shaped field of view rotates on the sky as the spacecraft is rotated about the line of sight to the star (asterisk), as illustrated by the numbered segments drawn with successively fainter outlines as the rotation sequence progresses in time. **b**, Three sample images at different rotation angles illustrate the observing sequence. Note that the planets (white '+' signs) are fixed in inertial space, and just barely visible. The rotation sequence continues to fill a full annular field of view. **c**, Roll deconvolution is applied to the data, removing the background speckles that rotate with the spacecraft, and keeping the part of the image (planets) fixed in the sky (see also Fig. 1 for relative intensities). The planets stand out clearly against the residual background noise, which is the time-varying part of the speckles.

Three snapshots of the planets-plus-speckle field are shown in Fig. 2b. With a relative speckle intensity of about 6×10^{-10} in each snapshot, the planets (centred under the white '+' signs) are barely visible. This combination of a slowly evolving instrumental background that is fixed with the rotating spacecraft, plus an astronomical object that is fixed in inertial space, has been encountered before with images from the HST. The technique of removing the background speckle field while retaining the astronomical objects in the image[20,21] is known as 'roll deconvolution'. We applied a roll deconvolution algorithm to the present sequence of laboratory background plus superposed planet images (J. Krist, personal communication). The result is shown in Fig. 2c, where we see that all three planets stand out clearly. Further, subtracting the azimuthally averaged background field yields an r.m.s. uncertainty of about 0.1×10^{-10}, as indicated by curve D in Fig. 1c. The Earth and the Jupiter intensities are shown added to the background from roll-deconvolution, and superposed for clarity.

The movie experiment is an existence proof that it is possible to extract exoplanets close to a star, even when the residual speckle intensity is comparable to, or greater than, the planet intensities. We believe, however, that for planet detection and characterization a space coronagraph should be designed to even stricter standards, with roughly a factor of ten weaker speckles than achieved in the present experiments. The present work is a step towards this goal, but more work remains. Further laboratory work now in progress is focused on pushing the speckle background lower, broadening the spectral bandwidth, suppressing speckles simultaneously on both sides of the star with a pair of DMs[22], and increasing the radial field of view, both inward and outward. The present work lays the groundwork for the development of future missions[2,23] that will, for the first time, explore nearby Earth-like exoplanet systems by direct imaging and spectroscopy.

Received 28 September 2006; accepted 26 February 2007.

1. Des Marais, D. J. *et al.* Remote sensing of planetary properties and biosignatures on extrasolar terrestrial planets. *Astrobiology* **2**, 153–181 (2002).
2. Brown, R. A. Single-visit photometric and obscurational completeness. *Astrophys. J.* **624**, 1010–1024 (2005).
3. Traub, W. A. *et al.* TPF-C: status and recent progress. *Proc. SPIE* **6268**, OT1–OT14 (2006).
4. Kaltenegger, L. & Fridlund, M. The Darwin mission: search for extra-solar planets. *Adv. Space Res.* **36**, 1114–1122 (2005).
5. Henry, C. *et al.* Terrestrial Planet Finder: architecture, mission design, and technology development. *Proc. SPIE* **5491**, 265–274 (2004).
6. Cash, W. Detection of Earth-like planets around nearby stars using a petal-shaped occulter. *Nature* **442**, 51–53 (2006).
7. Lyot, B. A study of the solar corona and prominences without eclipses. *Mon. Not. R. Astron. Soc.* **99**, 580–594 (1939).
8. Guyon, O., Pluzhnik, E. A., Kuchner, M. J., Collins, B. & Ridgway, S. T. Theoretical limits on extrasolar terrestrial planet detection with coronagraphs. *Astrophys. J. Suppl.* **167**, 81–99 (2006).
9. Born, M. & Wolf, E. *Principles of Optics* (Cambridge Univ. Press, New York, 2002).
10. Kuchner, M. J. & Traub, W. A. A coronagraph with a band-limited mask for finding terrestrial planets. *Astrophys. J.* **570**, 900–908 (2002).
11. Kuchner, M. J., Crepp, J. & Ge, J. Eighth-order image masks for terrestrial planet finding. *Astrophys. J.* **628**, 466–473 (2005).
12. Guyon, O. Phase induced amplitude apodization of telescope pupils for extrasolar terrestrial planet imaging. *Astron. Astrophys.* **404**, 379–387 (2003).
13. Kasdin, N. J., Vanderbei, R. J., Littman, M. J. & Spergel, D. N. Optimal one-dimensional apodizations and shaped pupils for planet-finding coronography. *Appl. Opt.* **44**, 1117–1128 (2005).
14. Vanderbei, R. J., Kasdin, N. J. & Spergel, D. N. Checkerboard-mask coronagraphs for high-contrast imaging. *Astrophys. J.* **615**, 555–561 (2004).
15. Soummer, R. Apodized pupil Lyot coronagraphs for arbitrary telescope apertures. *Astrophys. J.* **618**, L161–L164 (2005).
16. Riaud, P., Boccaletti, A., Baudrand, J. & Rouan, D. The four quadrant phase mask coronagraph. *Publ. Astron. Soc. Pacif.* **115**, 712–719 (2003).
17. Trauger, J. T. *et al.* Coronagraph contrast demonstrations with the high-contrast imaging testbed. *Proc. SPIE* **5487**, 1330–1336 (2004).
18. Trauger, J., Kern, B. & Kuhnert, A. *TPF-C Technology Milestone #1 Report* (JPL Document D-35484, Jet Propulsion Laboratory, Pasadena, 2006).
19. Borde, P. J. & Traub, W. A. High-contrast imaging from space: nulling in a low-aberration regime. *Astrophys. J.* **638**, 488–498 (2006).
20. Heap, S. R. *et al.* Space Telescope Imaging Spectrograph coronagraphic observations of β-Pictoris. *Astrophys. J.* **539**, 435–444 (2000).
21. Lowrence, P. J. *et al.* An infrared coronagraphic survey for substellar companions. *Astron. J.* **130**, 1845–1861 (2005).
22. Krist, J., Trauger, J. & Moody, D. Studying a simple TPF-C. *Proc. SPIE* **6265**, 3O1–3O11 (2006).
23. Trauger, J. *et al.* The Eclipse Mission: a direct imaging survey of nearby planetary systems. *Proc. SPIE* **4854**, 116–128 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The speckle nulling algorithm is due to C. Burrows. Development of the HCIT facility is an ongoing activity. Experiments were designed and simulated with the computational models of D. Moody. The HCIT control system was developed by B. Gordon and A. Niessner. Recent speckle nulling runs in white light have been carried out by B. Kern. The optical alignments were carried out by F. Shi and A. Kuhnert. The DM was manufactured by Xinetics Inc. The coronagraph mask was fabricated under the supervision of D. Wilson at JPL's Micro Devices Laboratory, on glass substrates supplied by Canyon Materials Inc. This work was carried out at JPL with the support of JPL and NASA.

Author Contributions J.T.T. performed HCIT experiments and data analysis at JPL; J.T.T. and W.A.T. co-wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.T.T. (john.trauger@jpl.nasa.gov).

LETTERS

Optical coherent state discrimination using a closed-loop quantum measurement

Robert L. Cook[1], Paul J. Martin[1] & J. M. Geremia[1]

Quantum mechanics hinders our ability to determine the state of a physical system in two ways: individual measurements provide only partial information about the observed system (because of Heisenberg uncertainty), and measurements are themselves invasive—meaning that little or no refinement is achieved by further observation of an already measured system[1]. Theoretical methods have been developed to maximize the information gained from a quantum measurement while also minimizing disturbance[2–4], but laboratory implementation of optimal measurement procedures is often difficult. The standard class of operations considered in quantum information theory[5] tends to rely on superposition-basis and entangled measurements[6], which require high-fidelity implementation to be effective in the laboratory[7]. Here we demonstrate that real-time quantum feedback[8–10] can be used in place of a delicate quantum superposition, often called a 'Schrödinger cat state', to implement an optimal quantum measurement for discriminating between optical coherent states[11,12]. Our procedure actively manipulates the target system during the measurement process, and uses quantum feedback to modify the statistics of an otherwise sub-optimal operator to emulate the optimal cat-state measurement. We verify a long-standing theoretical prediction[13] and demonstrate feedback-mediated quantum measurement[10,14] at its fundamental quantum limit over a non-trivial region of parameter space.

In this work we consider a single mode optical field $\hat{a}$ prepared at random into one of two pure coherent states, $|0\rangle$ or $|\alpha\rangle$, with a priori probabilities $P_0(0)$ and $P_0(\alpha)$. Privy only to the candidate states and their likelihoods, our problem is to determine whether the optical field was prepared into $|0\rangle$ or $|\alpha\rangle$ using the outcome of a measurement $\hat{M}$. On the basis of the theory of quantum hypothesis testing[15], it is sufficient to consider a two-component measurement, described by the operators $\hat{M}_0$ and $\hat{M}_\alpha = \hat{1} - \hat{M}_0$, with the interpretation that one selects $|0\rangle$ when the measurement outcome corresponds to $\hat{M}_0$ and vice versa. But because coherent states are not orthogonal, quantum mechanics precludes any measurement that can discriminate perfectly between $|0\rangle$ and $|\alpha\rangle$. The ability of a given operator $\hat{M}$ to distinguish between the two states in question can be characterized using (among other information theoretic measures) the probability of error[15,16]:

$$P_{\mathrm{E}} \equiv P_M(\alpha|0)P_0(0) + P_M(0|\alpha)P_0(\alpha) \tag{1}$$

Here, $P_M\left(\psi_{i\neq j}|\psi_j\right)$ is the conditional probability based on the measurement $\hat{M}$ that one erroneously selects $|\psi_i\rangle$ when the field was actually prepared into the state $|\psi_j\rangle$. $P_M(\psi_i|\psi_j)$ is determined by the quantum measurement statistics for $\hat{M}_i$ with respect to the state $|\psi_j\rangle$: specifically, $P_M(\alpha|0) = \mathrm{tr}\left[\hat{M}_\alpha|0\rangle\langle 0|\right]$ and $P_M(0|\alpha) = \mathrm{tr}\left[\hat{M}_0|\alpha\rangle\langle\alpha|\right]$. By weighting the two incorrect decisions by the a priori likelihoods of $|0\rangle$ and $|\alpha\rangle$, P_{E} quantifies the probability that one would mis-identify the field state in any individual measurement realization. We restrict ourselves here to equal a priori likelihoods, $P_0(0) = P_0(\alpha) = 1/2$, as this reflects the least classical prior information for use in decision making. Also, without loss of generality, our analysis applies to discriminating between any two coherent states $|\psi_0\rangle$ and $|\psi_1\rangle$ by performing an unconditional displacement to $|0\rangle$ and $|\alpha = \psi_1 - \psi_0\rangle$ before measurement.

Besides being a canonical problem in quantum optics, coherent state discrimination plays a fundamental role in communication theory[15]. For information to be transmitted between parties, the data must be encoded into the state of a physical system, which is itself subject to quantum mechanics. For example, in modern telecommunications, data are often encoded into an optical field using a discrete set of coherent state amplitudes by modulating the intensity or phase of a laser[11,13,16]. Achieving quantum limited discrimination between the different states of the laser field, and thus the communication symbols, is essential to minimizing the error rate when decoding the transmitted data[12,17]. Distinguishing between weak coherent states is also important to applications of quantum information science, in particular for quantum key distribution.

To determine the optimal measurement for discriminating between coherent states, one minimizes equation (1) over all valid quantum maps. Doing so leads to the quantum limit[15,16]:

$$P_{\mathrm{QL}} = \frac{1}{2}\left(1 - \sqrt{1 - \mathrm{e}^{-\bar{n}_\alpha}}\right) \tag{2}$$

achieved by the so-called 'cat-state' measurement: $\hat{M}_0^* = |m\rangle\langle m|$ where $|m\rangle = c_0(\alpha)|0\rangle + c_1(\alpha)|\alpha\rangle$ (the optimal superposition depends on α)[15]. Unfortunately, preparing such a cat state in the laboratory is generally considered extremely difficult[17,18], so much so that it has not been accomplished in the four decades since this problem was first studied.

The standard sub-optimal approach to coherent state discrimination resorts to measuring the number operator, $\hat{n} = \hat{a}^\dagger\hat{a}$, implemented approximately in the laboratory by direct photon counting[15]. State discrimination is performed by partitioning the number operator outcomes into two projectors, $\hat{M}_0 = |0\rangle\langle 0|$ and $\hat{M}_\alpha = \hat{1} - |0\rangle\langle 0|$, because (technical imperfections aside) photon counting produces $n > 0$ only when the field was prepared into $|\alpha\rangle$. An error occurs when no photons are detected due to vacuum fluctuations in $|\alpha\rangle$. The resulting state discrimination error is known as the shot-noise error probability:

$$P_{\mathrm{SN}} = \mathrm{e}^{-\bar{n}_\alpha}/2 \tag{3}$$

where $\bar{n}_\alpha = |\alpha|^2$ is the mean photon number for $|\alpha\rangle$.

In this work we demonstrate that one can surpass shot noise and even approach the quantum limit using real-time quantum feedback[8–10] in place of the cat-state measurement. Our approach exploits the finite duration of any real measurement. In the present context, the quantum states $|0\rangle$ and $|\alpha\rangle$ are realized as optical wavepackets with spatiotemporal extent. Measurements on an optical pulse inherently

[1]Quantum Measurement & Control Group, Department of Physics and Astronomy, The University of New Mexico, Albuquerque, New Mexico 87131, USA.

persist for a time set by the pulse length, τ. For example, photon counting generates a measurement record $\Xi_{[0,\tau]} = (t_1, t_2, \ldots, t_n)$ that consists of the observed photon arrival times even if $\hat{M}$ is modelled using standard quantum measurement theory by viewing the total number of photon arrivals in the counting interval $[0,\tau]$ as one aggregate 'instantaneous' measurement of the number operator. Rather, we exploit this time to feed back on the optical field in a manner that emulates the actual cat-state measurement.

A diagram of our closed-loop measurement is shown in Fig. 1a. Following the proposal by Dolinar[13], we combine photon counting with feedback-mediated optical displacements applied during the photon counting interval. The amplitude of the displacement u_t applied at each time t during the measurement is conditional on the accumulated measurement record $\Xi_{[0,t]}$ and based on an evolving bayesian estimate of the incoming wavepacket state[13,17]. Discrimination is performed by selecting the state $|\psi\rangle \in \{|0\rangle, |\alpha|\}$ that maximizes the conditional probability $P(\Xi_{[0,t]}|\psi, u_{[0,t]})$ that the measurement record $\Xi_{[0,t]}$ would be observed given the state $|\psi\rangle$ and the history of applied displacements $u_{[0,t]}$. The feedback controller determines which state is most consistent with the accumulating record $\Xi_{[0,t]}$ and chooses the feedback amplitude at each point in time to minimize the probability of error over the remainder of the measurement interval $(t, \tau]$. The policy for determining the optimal displacement amplitude u_t^* is found by minimizing a time-additive extension of equation (1)[17]:

$$P_E[u_t] = \frac{1}{2}\int_0^\tau dt \left[P(\alpha|0, u_{[0,t]}) + P(0|\alpha, u_{[0,t]})\right] \quad (4)$$

that demands minimal error at all times during the measurement, not just in its aggregate statistics. Here, $P(\psi_i|\psi_j, u_{[0,t]})$ is the probability of selecting $|\psi_i\rangle$ given the field state $|\psi_j\rangle$ and the displacement history $u_{[0,t]}$ averaged over measurement records. Functional minimization of equation (4) was first performed by Dolinar[13] and more recently using modern optimal control theory[17]. The resulting feedback policy:

$$u_t^*(n_{[0,t)}) = \frac{\alpha}{2}\left(\frac{e^{i\pi(n_{[0,t)}+1)}}{\sqrt{1-e^{-\bar{n}_\alpha t/\tau}}} - 1\right) \quad (5)$$

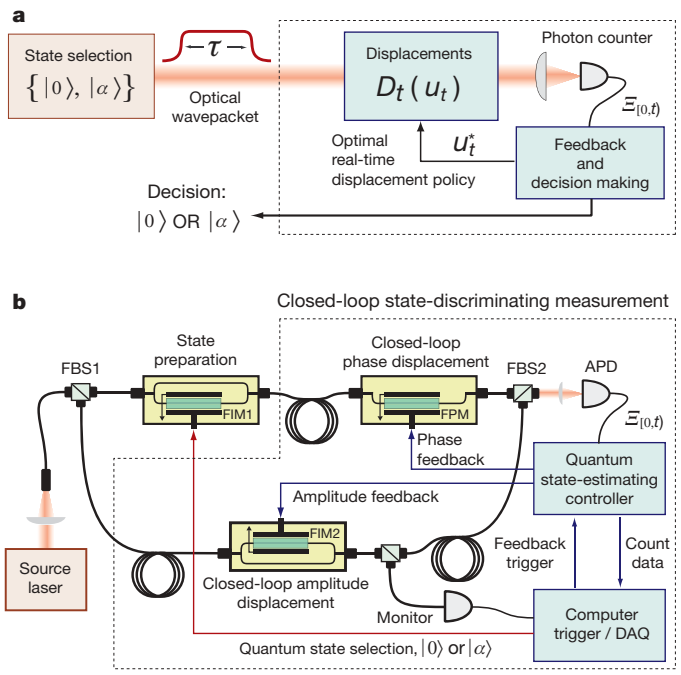


Figure 1 | Closed-loop quantum state discrimination. **a**, We consider a measurement that combines photon counting with feedback-mediated optical displacements to enact quantum-limited state discrimination between the coherent states $|0\rangle$ and $|\alpha\rangle$. **b**, A simplified diagram of our laboratory implementation of **a**. See main text for details.

with the decision procedure that $|\alpha\rangle$ ($|0\rangle$) is chosen when the number of photon counts in the measurement interval $n_{[0,\tau]}$ is even (odd), analytically achieves the fundamental quantum limit, equation (2)[13,17].

Figure 1b shows a diagram of our laboratory implementation of the closed-loop measurement just described. Light from an external-cavity grating-stabilized diode laser operating at 852 nm is coupled into a polarization maintaining (PM) fibre-optic Mach–Zender interferometer. The input beamsplitter (FBS1) provides two optical fields with a well-defined relative phase: the upper arm of the interferometer acts as the target quantum system for state discrimination and the lower arm provides an auxiliary field used to perform closed-loop displacements at the second beamsplitter (FBS2). Photon counting on the outcoupled field is implemented using a gated silicon avalanche photodiode (APD), and our feedback controller is constructed from a combination of programmable waveform generators and high-speed digital signal processing electronics (feedback bandwidth ~30 MHz). A digital counter records the number of photon counter clicks generated in each measurement interval

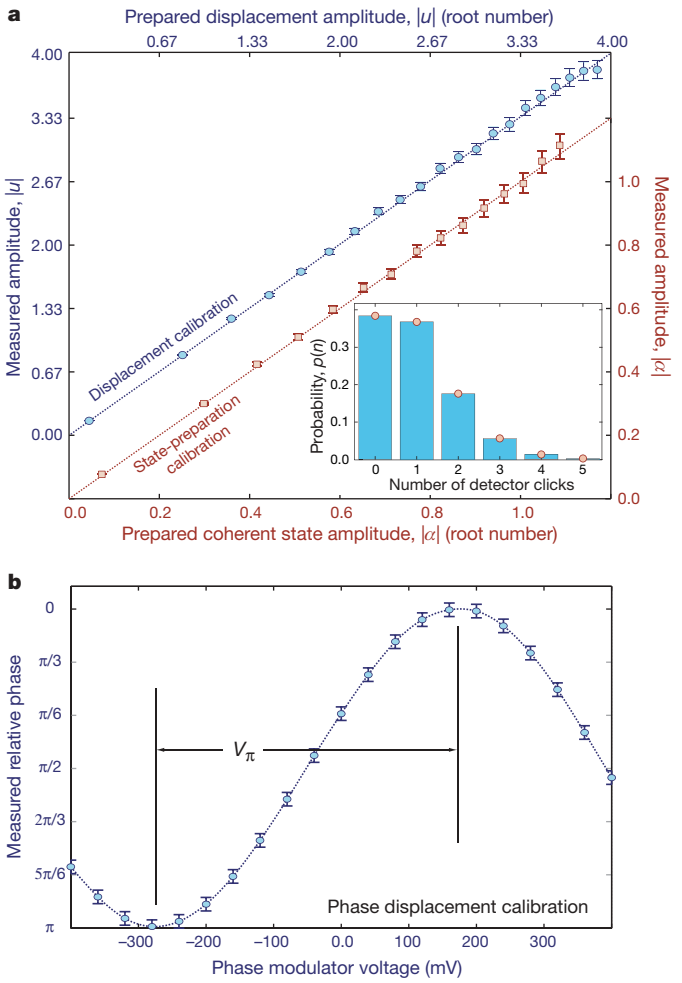


Figure 2 | State preparation and control displacement calibration. **a**, Demonstration of our ability to prepare optical coherent states (blue circles) and perform controlled displacements (red squares) with high fidelity using a precision calibration of our apparatus. Inset, a Poisson fit of our photon counting statistics for $\alpha \approx 1$, illustrating that our state preparation is quantum-limited. **b**, Demonstration of our ability to apply controlled phase displacements $\phi(u_t)$. The calibration in **b** illustrates the low control voltages (V_π) required to drive our apparatus, which allows for high-bandwidth application of the measurement feedback control. Each data point in **a** and **b** reflects a statistical ensemble of 100,000 replicate measurements, with error bars given by the estimated sample standard deviation.

$[0,\tau]$, during which the feedback controller determines the feedback amplitude, equation (5), via the accumulating count record $n_{[0,t)}$.

Coherent states for discrimination are realized as $\tau = 20\,\mu s$ optical pulses produced by a computer-controlled PM fibre-optic intensity modulator (FIM1) in the upper arm of the interferometer. The calibration parity between desired and observed values of $|\alpha|$ in Fig. 2a (red squares) highlights our ability to prepare arbitrary optical coherent states with amplitudes $0.1 \leq |\alpha| \leq 1$ (the phase of α is described below). Counting statistics for one such preparation with $\bar{n}_\alpha \approx 1$ is shown in the inset plot. The red circles in the inset are a poissonian fit to the counting data, for which we compute $\chi^2 - 1$ below 1 p.p.m., providing compelling evidence that our state preparation is quantum noise-limited. We perform a similar calibration analysis using the intensity modulator in the lower arm of the interferometer (FIM2). The amplitude in this arm is used to implement the magnitude of the control displacement $|u_t|$ applied to the target quantum state via the outcoupling beamsplitter FBS2. Figure 2a plots the parity between the desired and observed displacement amplitude $|u_t|$ (blue circles). The phase of our prepared coherent states $\phi(\alpha)$ and the phase of the feedback displacements $\phi(u_t)$ are implemented by the modulator (FPM) in the upper arm of the interferometer. Without loss of generality, we always chose $\phi(\alpha) = 0$ to simplify the interpretation of our displacements. Figure 2b illustrates our ability to control the relative interferometer phase and thus apply $\phi(u_t)$. We actively stabilize the laser intensity, interferometer path length, and modulator temperature to enable accurate comparison of different statistical ensembles. Residual technical imperfections in our experiment result primarily from detector dark counts ($\bar{n}_d = 0.0078$), interferometer phase noise ($\delta\phi \approx 8$ mrad), and finite extinction of our modulators.

With feedback disabled, our measurement reduces to direct photon counting. The data (red squares) in Fig. 3a show that, in the absence of feedback, our observed probability of error for discriminating between $|0\rangle$ and $|\alpha\rangle$ faithfully reproduces shot noise (solid red line) as a function of $\bar{n}_\alpha = |\alpha|^2$. Each data point was calculated using 100,000 optical pulses sampled randomly from $\{|0\rangle, |\alpha\rangle\}$ with equal probability. We use the label $|0\rangle$ to signify the darkest field $\bar{n}_0 \approx 0.008$ we can achieve given finite intensity-modulator extinction. The residual field appears to have a negligible effect, with a discrepancy between our photon counting data in Fig. 3a and equation (3) of $\chi^2 - 1 = 1.13 \times 10^{-5}$.

The data in Fig. 3a have been adjusted to account for finite detector efficiency and optical losses. That is to say, the coherent state amplitudes in our experiment are normalized with respect to the average photon number $\bar{n}_\alpha = |\alpha|^2$ measured by the APD over a $t = 20\,\mu s$ square-envelope pulse. Owing to the nature of coherent states, it has been shown that the detection efficiency η (resulting from the combination of detector quantum efficiency η_d and optical efficiency η_e) factors out of a comparison between the shot-noise and quantum limts[17]. For comparison, we have plotted the shot-noise error and quantum limits that would correspond to ideal detection ($\eta = 1$) in Fig. 3a. We have independently determined the intrinsic efficiency of our apparatus to be approximately $\eta \approx 0.35$.

Figure 3b, traces 1–4, shows an example closed-loop measurement trajectory in which the field was prepared into the state $|\alpha = 0.20\rangle$, to be distinguished from $|0\rangle$. The measured pulse envelope of the prepared coherent state is depicted in trace 1 and the shaded regions ($t < 0$ and $t > \imath$) indicate that the APD is gated off at those times. Before $t = 0$, the displacement amplitude is pre-established to its initial value to suppress ringing and slew-rate limitations. Careful inspection of equation (5) reveals that the optimal displacement diverges at early time, $|u_0| \to \infty$. However, finite modulator extinction (and APD saturation to a lesser extent) limits the practical maximum displacement that can be applied in trace 2. Our maximum displacement corresponds to $\bar{n}_u \approx 15.8$, although this could be increased with additional intensity modulators in future experiments.

Many aspects of the closed-loop measurement are evident from the single-shot trajectory in Fig. 3b. At $t = 0$ there is no reason to prefer one state, $|0\rangle$ or $|\alpha\rangle$, over the other. But as more data become available, the controller refines its bayesian estimate of the incoming optical state by updating the conditional probabilities $P\left(\psi|\Xi_{[0,t]}, u_{[0,t]}\right)$, described above. The premiss behind the closed-loop measurement is to displace the field to the vacuum in each shot and decide which state is present based on the displacement applied to cancel the field. As the controller gains increased confidence in its guess, it is better able to perform the correct nulling displacement. From equation (5),

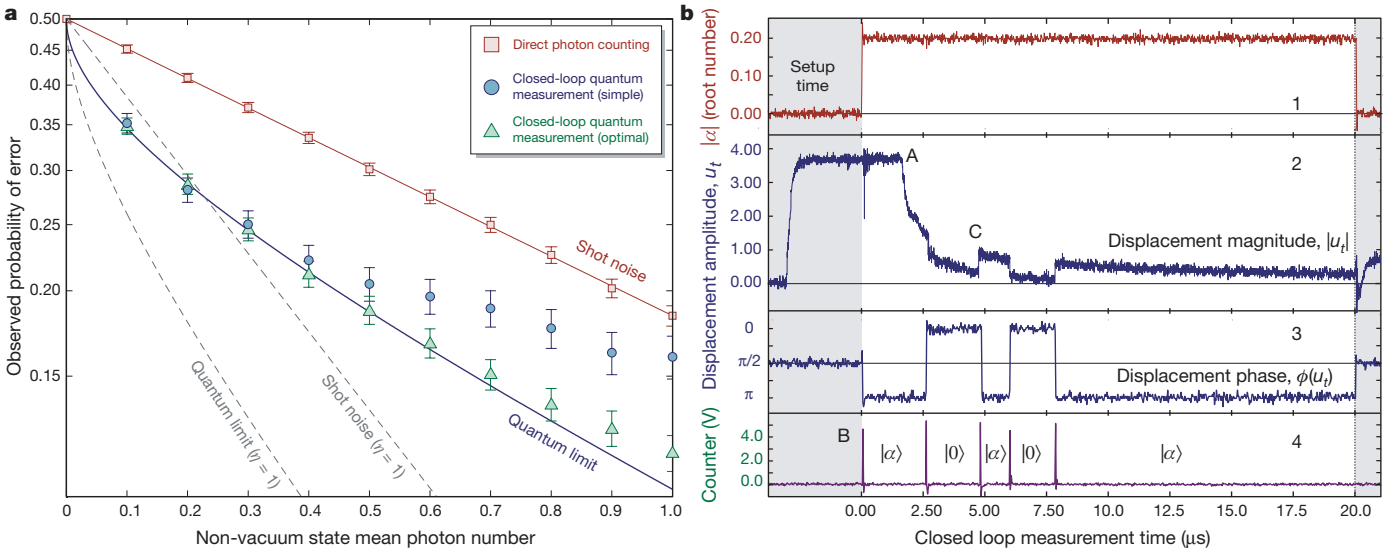


Figure 3 | Laboratory observation of sub-shot-noise state discrimination. **a**, The measured probability of error versus mean photon number for both direct photon counting (red squares) and our closed-loop measurement interpreted using a bayesian estimator that assumes application of the optimal closed-loop control policy (blue circles) and one that accounts for experimental imperfections (green triangles). All data points were obtained from ensembles of 100,000 measurement trajectories, with error bars that reflect the sample standard deviation. The four traces in **b** (1–4) depict a single-shot closed-loop measurement trajectory, discussed in detail in the text. We draw attention to several technical issues: first, the finite dynamical range of our displacements (point A on graph); second, the initial APD click is a timing signal, not a real detection event (B); and third, the apparent rise time is that of the monitor photodiode not the feedback (C).

the displacement magnitude $|u_t^*|$ is inversely proportional to the time-dependent decision uncertainty $\sqrt{1-\exp(-\bar{n}_\alpha t/\tau)}$; although not readily apparent, performing the optimization of equation (4) reveals that it is statistically optimal for the closed-loop measurement to reverse its state hypothesis with each detector click during the counting interval[13,17]. The sequence of hypothesis reversals in the example closed-loop trajectory is denoted along with the measurement record in Fig. 3b trace 4. As the measurement record accumulates, the controller eventually settles on its final (correct) decision, which in this case is $|\alpha\rangle$.

The data (blue circles) in Fig. 3a demonstrate that the closed-loop state discrimination procedure described above (alternating guesses between $|\alpha\rangle$ and $|0\rangle$ with each photon arrival) surpasses the shot-noise error probability for amplitudes $|\alpha|$ less than about one. We essentially saturate the fundamental quantum limit over a non-trivial region of parameter space $\bar{n}_\alpha < 0.3$. Performance does degrade at larger mean photon numbers owing to technical limitations: most significantly, finite extinction in the intensity modulator prevents $|u_t|$ from achieving the optimal magnitude $|u_t^*|$ at early measurement times—the contrast ratio of the modulator prevents us from inputting an arbitrarily large field into the lower arm of our interferometer (see Fig. 1) as the closed-loop displacement policy for $|0\rangle$ requires $|u_t| \approx 0$.

The data (blue circles) in Fig. 3a were determined assuming that we implemented the optimal feedback control policy in equation (5) perfectly by selecting the state $|\psi\rangle \in \{|0\rangle, |\alpha\rangle\}$ to maximize the conditional probability $P\left(\Xi_{[0,t]}|\psi, u^*_{[0,\tau]}\right)$. This approach is clearly sub-optimal owing to the technical imperfections in our control displacements just described. On the basis of a suggestion by H. M. Wiseman (personal communication; a similar approach was taken in the adaptive phase measurement)[14], we reinterpreted our raw measurement data to account for deviations between the feedback actually performed in our experiment $u_{[0,\tau]}$ and the optimum policy $u^*_{[0,\tau]}$. The data (green triangles) in Fig. 3a reflect an analysis based on the true conditional probability $P\left(\Xi_{[0,t]}|\psi, u_{[0,\tau]}\right)$, and we find that our procedure nearly achieves the quantum limit (for our detection efficiency) over the full range of coherent states we investigated. This improvement to Dolinar's original solution[13] was made possible by an optimal control theoretic treatment of coherent state discrimination[17]. We observe that even with detection efficiency $\eta \approx 0.35$ our closed-loop measurement slightly outperforms the ideal shot-noise error that would be achieved in a technically lossless experiment $\eta = 1$ for photon numbers $\bar{n} < 0.2$.

Our experiment complements previous work on adaptive quantum measurements[14,19]: rather than optimizing over a parameterized space of measurements (for example, quadrature operators) for a fixed state, our procedure manipulates the state of the observed system for a fixed measurement. Quantum feedback can be viewed as manipulating the outcome statistics of the number operator $\hat{n}$. In the absence of feedback, the detailed measurement record consisting of photon arrival times $\Xi_{[0,\tau]} = (t_1, t_2, \ldots, t_n)$ provides no more information than the total number n: Poisson processes are stationary in time, but with feedback, the significance of each click depends on when it occurs, even though the field is described by some coherent state at each point in time. The optimal feedback policy applies displacements in a manner that extracts as much information out of each photon arrival as possible. It is in this manner that we surpass shot noise to achieve the fundamental quantum limit over a non-trivial range of $|\alpha|$.

Furthermore, our procedure is evidently less demanding on the measurement resources needed to achieve optimal statistics than a direct implementation of a cat state. At no point in time do we generate a superposition between optical coherent states, yet we effectively achieve the optimal result by exploiting the time-dependence of the measurement. Our results suggest that demanding quantum resources such as entanglement and superposition states may in fact be fungible. At least here, it was possible to replace the use of such fragile resources with a more robust method—quantum feedback.

Received 14 December 2006; accepted 1 February 2007.

1. von Neumann, J. *Mathematical Foundations of Quantum Mechanics* (Princeton Univ. Press, Princeton, 1955).
2. Kraus, K. *Lecture Notes in Physics* Vol. 190, *States, Effects, and Operations: Fundamental Notions of Quantum Theory* (Springer, Berlin, 1983).
3. Peres, A. & Wooters, W. K. Optimal detection of quantum information. *Phys. Rev. Lett.* **66,** 1119–1122 (1991).
4. Fuchs, C. A. & Peres, A. Quantum-state disturbance versus information gain: Uncertainty relations for quantum information. *Phys. Rev. A* **53,** 2038–2045 (1996).
5. Nielsen, M. & Chuang, I. *Quantum Computation and Quantum Information Science* (Cambridge Univ. Press, Cambridge, UK, 2000).
6. Peres, A. Neumark's theorem and quantum inseparability. *Found. Phys.* **20,** 1441–1453 (1990).
7. Leibfried, D. *et al.* Toward Heisenberg-limited spectroscopy with multiparticle entangled states. *Science* **304,** 1476–1478 (2004).
8. Wiseman, H. M. & Milburn, G. J. All-optical versus electro-optical quantum-limited feedback. *Phys. Rev. A* **49,** 4110–4125 (1994).
9. Doherty, A. C., Habib, S., Jacobs, K., Mabuchi, H. & Tan, S. M. Quantum feedback control and classical control theory. *Phys. Rev. A* **62,** 012105–012117 (2000).
10. Geremia, J., Stockton, J. K. & Mabuchi, H. Real-time quantum feedback control of atomic spin-squeezing. *Science* **304,** 270–273 (2004).
11. Helstrom, C. W. Detection theory and quantum mechanics. *Inform. Control* **10,** 254–291 (1967).
12. Sasaki-Usuda, T. & Hirota, O. in *Quantum Communication and Measurement* (eds Belavkin, V. P., Hirota, O. & Hudson, L.) 419–428 (Plenum, New York, 1995).
13. Dolinar, S. *Quarterly Progress Report* (Tech. Rep. 111, Research Laboratory of Electronics, MIT, 1973).
14. Armen, M. A., Au, J. K., Stockton, J. K., Doherty, A. C. & Mabuchi, H. Adaptive homodyne measurement of optical phase. *Phys. Rev. Lett.* **89,** 133602–133605 (2002).
15. Helstrom, C. W. *Mathematics in Science and Engineering* Vol. 123, *Quantum Detection and Estimation Theory* (Academic, New York, 1976).
16. Fuchs, C. A. *Distinguishability and Accessible Information in Quantum Theory*. Ph.D. thesis, Univ. New Mexico (1996); ⟨http://arxiv.org/quant-ph/9601020⟩ (1996).
17. Geremia, J. Distinguishing between optical coherent states with imperfect detection. *Phys. Rev. A* **70,** 062303–062311 (2004).
18. Sasaki, M. & Hirota, O. Quantum decision scheme with a unitary control process for binary quantum-state signals. *Phys. Rev. A* **54,** 2728–2736 (1996).
19. Wiseman, H. M. Adaptive phase measurements of optical modes: Going beyond the marginal Q distribution. *Phys. Rev. Lett.* **75,** 4587–4590 (1995).

Acknowledgements We thank H. M. Wiseman, H. Mabuchi, S. Dolinar, K. Jacobs, A. Landahl, C. Caves, I. Deutsch and P. Jessen for discussions. This work was supported by the AFOSR and the New Mexico NSF EPSCoR programme in nanotechnology.

Author Contributions All authors contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.M.G. (jgeremia@unm.edu).

Vol 446|12 April 2007|doi:10.1038/nature05669

LETTERS

Rapid and reversible shape changes of molecular crystals on photoirradiation

Seiya Kobatake[1], Shizuka Takami[2], Hiroaki Muto[2], Tomoyuki Ishikawa[1] & Masahiro Irie[2]

The development of actuators based on materials that reversibly change shape and/or size in response to external stimuli has attracted interest for some time[1]. A particularly intriguing possibility is offered by light-responsive materials, which allow remote operation without the need for direct contact to the actuator. The photo-response of these materials is based on the photoisomerization of constituent molecules (typically *trans–cis* isomerization of azobenzene chromophores), which gives rise to molecular motions and thereby deforms the bulk material. This effect has been used to create light-deformable polymer films and gels[2–10], but the response of these systems is relatively slow. Here we report that molecular crystals based on diarylethene chromophores and with sizes ranging from 10 to 100 micrometres exhibit rapid and reversible macroscopic changes in shape and size induced by ultraviolet and visible light. We find that on exposure to ultraviolet light, a single crystal of 1,2-bis(2-ethyl-5-phenyl-3-thienyl)perfluorocyclopentene changes from a square shape to a lozenge shape, whereas a rectangular single crystal of 1,2-bis(5-methyl-2-phenyl-4-thiazolyl)perfluorocyclopentene contracts by about 5–7 per cent. The deformed crystals are thermally stable, and switch back to their original state on irradiation with visible light. We find that our crystals respond in about 25 microseconds (that is, about five orders of magnitude faster than the response time of the azobenzene-based polymer systems[7–10]) and that they can move microscopic objects, making them promising materials for possible light-driven actuator applications.

Azobenzene-containing liquid-crystal elastomers have been shown[7] to contract by as much as 22% on ultraviolet irradiation. The effect can give rise to directed bending of elastomer films, provided the chromophores are selectively excited with linearly polarized light[8] or unidirectionally aligned in the film by a rubbing procedure[9,10]. In both cases, the light-induced *trans–cis* photoisomerization of the azobenzene chromophores reduces the ordering of the liquid-crystal material, which in turn gives rise to macroscopic contraction or bending. The response time of these systems is, however, rather slow (typically on the timescale of seconds to minutes), and the deformed states are unstable because the *cis*-azobenzene isomers relax back to *trans*-azobenzenes[7]. In contrast, diarylethene molecules undergo thermally irreversible and fatigue-resistant photochromic reactions in solution as well as in the single-crystalline phase[11–13]. The effect has been used[14] to reversibly change the surface morphology of a diarylethene single crystal on photoirradiation.

Figure 1 shows 1,2-bis(2-ethyl-5-phenyl-3-thienyl)perfluorocyclopentene (**1**) and 1,2-bis(5-methyl-2-phenyl-4-thiazolyl)perfluorocyclopentene (**2**) used in this study, and illustrates their structural transformations and single-crystal deformations on alternate irradiation with ultraviolet (wavelength 365 nm) and visible (>500 nm) light (see Supplementary Videos 1 and 2). Single crystals of **1** and **2**, with sizes on the 10–100 μm scale, were prepared by sublimation of the compounds on glass plates. On irradiation with ultraviolet light, the molecules in the crystals underwent a cyclization reaction that transforms open-ring isomers into closed-ring isomers and changes the colours of crystals **1** and **2** to blue and violet, respectively. The colours were stable in the dark, but disappeared on

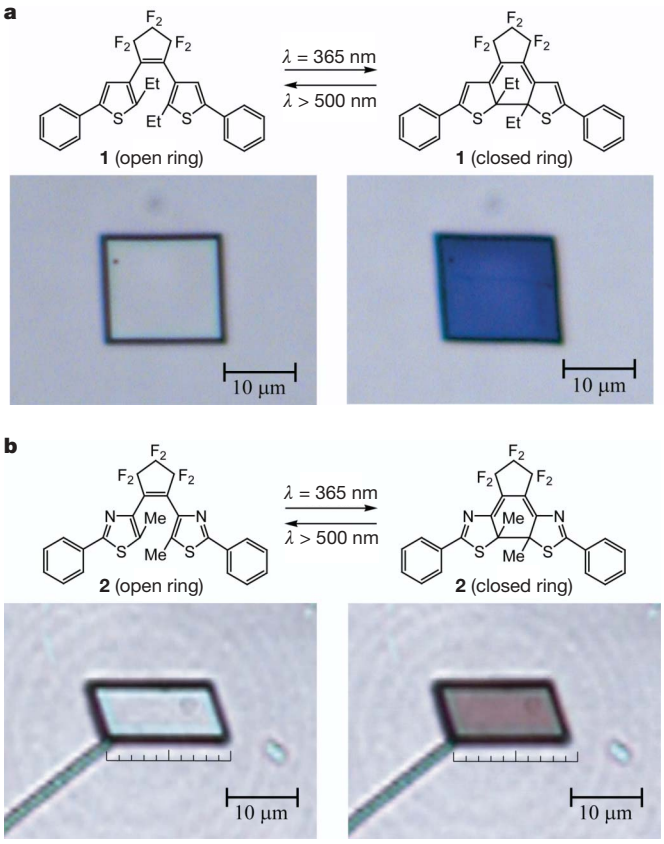


Figure 1 | Chemical structures and deformation of the two diarylethene compounds. Molecular structures of the open- and closed-ring isomers of the two compounds along with images illustrating the deformation of their single crystals on irradiation with ultraviolet (365 nm) and visible (>500 nm) light. **a**, A square single crystal of 1,2-bis(2-ethyl-5-phenyl-3-thienyl)perfluorocyclopentene (**1**) with corner angles of 88° and 92° reversibly changed to a lozenge shape with corner angles of 82° and 98°. The thickness of the crystal was 570 nm. **b**, A rectangular single crystal of 1,2-bis(5-methyl-2-phenyl-4-thiazolyl)perfluorocyclopentene (**2**) reversibly contracted and expanded by as much as 7%. The thickness of the crystal was 330 nm.

[1]Department of Applied Chemistry, Graduate School of Engineering, Osaka City University, Sugimoto 3-3-138, Sumiyoshi-ku, Osaka 558-8585, Japan. [2]Department of Chemistry and Biochemistry, Graduate School of Engineering, Kyushu University, Hakozaki 6-10-1, Fukuoka 812-8581, Japan.

irradiation with visible light—as expected for photochemical isomerization back to the initial open-ring isomers[15].

Ultraviolet irradiation of a single crystal of **1** changed its corner angles from 88° and 92° to 82° and 98°, respectively, and hence its shape from a square to a lozenge (or from a hexagon to a deformed hexagon, as shown in Supplementary Video 1). Figure 2a documents the time dependence of the colour and shape changes on alternate irradiation with ultraviolet and visible light by showing the relation between the absorption intensity of the crystal at 600 nm wavelength and its corner angle. We note that the absorption intensity increases with the amount of photogenerated, coloured, closed-ring isomers present in the crystal, which reaches 70% of all molecules at the photostationary state. (See Supplementary Fig. S1, which documents infrared absorption measurements on a crystal of **1** with a thickness less than 1 μm, under irradiation with 365 nm light.) The data in Fig. 2a show that the angle initially remains unchanged and then decreases by as much as 5° to 6°. We found no evidence for hysteresis between the forward and reverse processes, but observed an interesting correlation between the absorption maximum and the shape change: the absorption maximum initially remained constant at 625 nm, but then shifted to 585 nm as the crystal shape changed (see Supplementary Fig. S2). The spectral shift is attributable to the formation of adjacent closed-ring isomers[16]. We infer from this correlation between the onset of the shape change and of the spectral shift that intermolecular interactions between adjacent closed-ring isomers induce the shape change.

To evaluate the crystallinity of the small crystal used in the switching experiment, its melting point and the order parameter of the visible absorption were measured at the photostationary state. The melting point of crystal **1** before photoirradiation was 164 °C, decreased to 45–55 °C on irradiation with ultraviolet light, and returned to 164 °C on irradiation with visible light. The melting point of 164 °C is the same as observed with a large crystal. The order parameter, $(A_{//} - A_{\perp}) / (A_{//} + 2A_{\perp})$, at 600 nm at the photostationary state was 0.53, which is identical to the value measured with a large crystal[15]. (Here $A_{//}$ is the absorbance parallel to the long axis of the molecules packed in the crystal and $A_{\perp}$ is the absorbance normal to the axis, measured with linearly polarized light.) The recovery of the melting point on visible-light irradiation indicates that crystal **1** remains highly crystalline after a cycle of irradiation with ultraviolet and visible light. The decrease of the melting point on ultraviolet irradiation is due to the coexistence of two isomers in the same crystal. The reversibility of the angle changes was also examined, with no evidence for a change in the performance of the crystal—even after 20 cycles of alternate irradiation with ultraviolet and visible light (see Supplementary Fig. S3).

As shown in Fig. 1b, irradiation of a rectangular single crystal of **2** with ultraviolet and visible light induces contraction and expansion, respectively, by as much as 7%. The correlation between the absorption intensity of the closed-ring isomers at 550 nm and changes in the

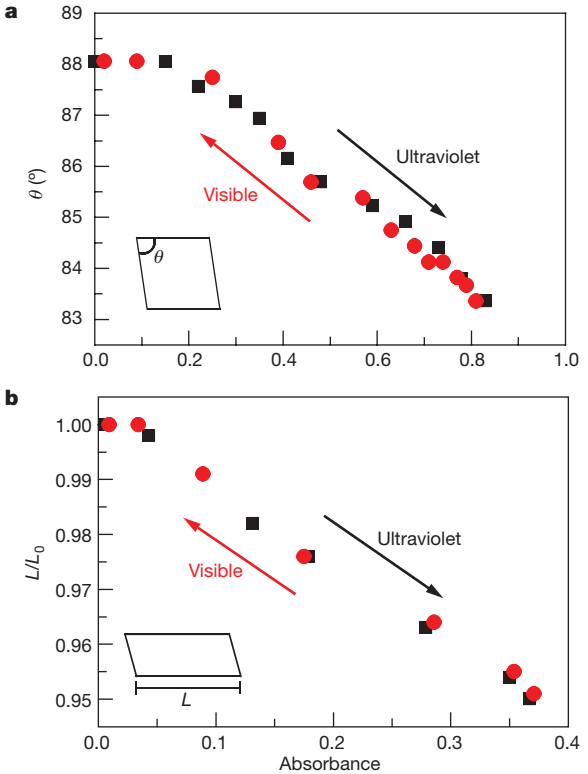


Figure 2 | Time dependence of the photo-response of single crystals of 1 and 2. a, Relationship between the corner angle of the single crystal of **1** and the absorption intensity of the crystal measured at 600 nm on alternate irradiation with ultraviolet (365 nm, filled black squares) and visible (>500 nm, filled red circles) light. **b**, Relationship between the length of the rectangular single crystal of **2** and the absorbance of the crystal measured at 550 nm on alternate irradiation with ultraviolet (365 nm, filled black squares) and visible (>500 nm, filled red circles) light. L_0, original length; L, length after irradiation.

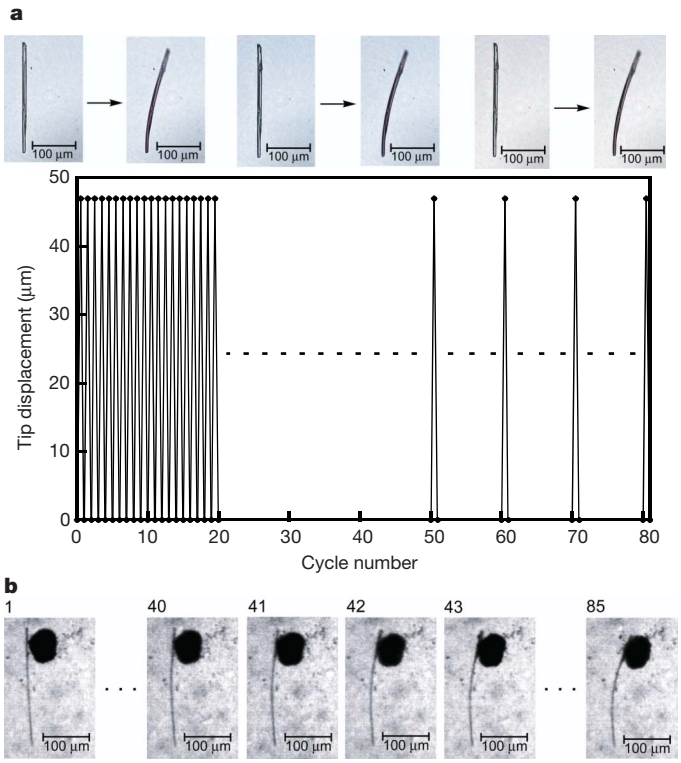


Figure 3 | Reversible bending of a crystalline rod. a, Reversible bending of a rod-like crystal of **2** on alternate irradiation with ultraviolet (365 nm) light for 0.1 s and visible (>500 nm) light for 6 min (main panel). On irradiation with ultraviolet light, the rod-like crystal (225 × 7.5 × 5 μm) bent and the head moved as far as 47 μm. The bent rod-like crystal straightened again on irradiation with visible light. The reversible bending could be repeated for as many as 80 cycles. Top panel, pairs of images showing the first, fiftieth and eightieth cycles (left to right). **b**, Movement of a gold micro-particle by the rod-like crystal on irradiation with ultraviolet (365 nm) light. The gold micro-particle is 90 times heavier than the rod-like crystal (250 × 5 × 5 μm) and appears in the images as a black spot. The rod-like crystal bent when it was irradiated with ultraviolet light and could push the gold micro-particle as far as 30 μm. Photographs were taken with a high-speed camera (Phantom V7.2, Vision research) with an image intensifier (C6653MOD, Hamamatsu Photonics). The exposure time of each frame was 500 μs (2,000 frames sec^{-1}). The numbers above the images are frame numbers.

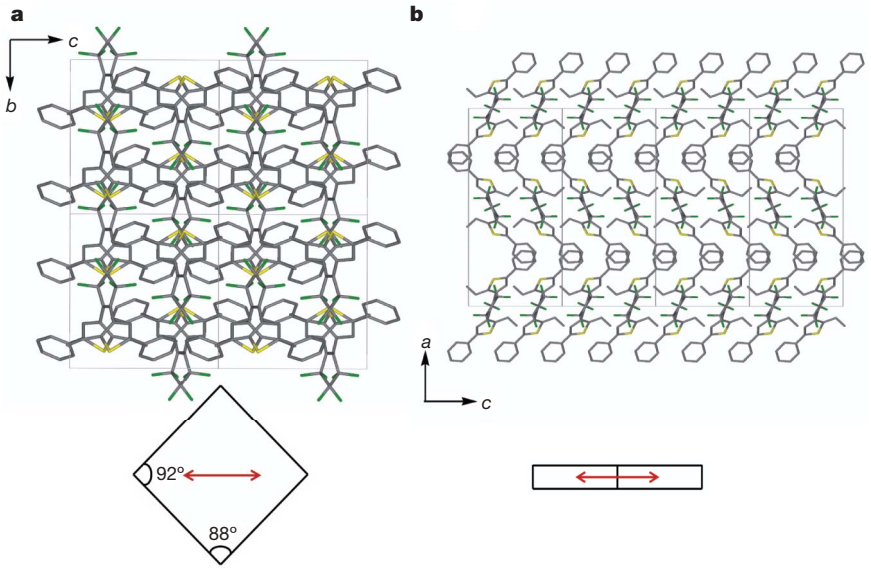


Figure 4 | Molecular packing of a crystal of 1. Crystalline packing of molecules of **1** in the crystal before ultraviolet irradiation and outlines of the crystal morphology as viewed from **a**, the (100) face, and **b**, the (010) face. Red arrows in diagrams below show the direction of contraction of the crystal (*c* axis).

length of the crystal is shown in Fig. 2b. Just as observed with crystal **1**, after an induction period during the initial stage of ultraviolet irradiation, the crystal length linearly decreases by about 5–7%. We found no evidence for any hysteresis between forward and reverse processes, and were able to carry out more than 20 cycles of light-driven contraction and expansion without evidence for any appreciable fatigue. The initial melting point of the crystal decreased from 158 °C to 95–105 °C on ultraviolet irradiation and returned to 158 °C on irradiation with visible light, whereas the order parameter of the absorption at 550 nm at the photostationary state was as high as 0.75. These observations indicate that the crystal of **2** also remains highly crystalline after a cycle of irradiation with ultraviolet and visible light.

We also prepared rod-like crystals of **2** (Fig. 3). X-ray crystallographic analysis revealed that the thin, rectangular plate-like crystal and the rod-like crystal have the same crystal structure. The rod-like crystal bent on irradiation with ultraviolet light (see Supplementary Video 3), with the bending moving towards the direction of the incident light; one end of the rod was fixed to the glass plate. We ascribe this effect to a gradient in the extent of photoisomerization caused by the high absorbance of the crystal, so that the shrinkage of the irradiated part of the crystal causes bending—just as in a bimetal. The bent rod-like crystal became straight again on irradiation with visible light. Such reversible bending on alternate irradiation with ultraviolet and visible light could be repeated over 80 cycles (Fig. 3a). The power produced during bending is illustrated in Fig. 3b, which shows how the photoinduced bending of a rod-like crystal ($250 \times 5 \times 5$ μm) moves a gold micro-particle with a weight 90 times greater than the single crystal over a distance of 30 μm (see also Supplementary Videos 4 and 5). As before, the rod-like crystal straightened again on irradiation with visible light.

The photoreaction of diarylethenes in crystals takes place in less than 10 ps (ref. 17), but it is not clear how this timescale relates to that of the mechanical changes of the crystal. To reveal the dynamics of the crystal bending, a high-speed camera with an image intensifier was used to measure the bending rate after irradiation with pulsed laser light (wavelength 355 nm; pulse width, 8 ns; power, 60 mJ per pulse). Following irradiation with a single 8-ns-long pulse, the bending of the crystal was found to be almost complete within 25 μs (see Supplementary Fig. S4). This very fast response time is about 10^5 times faster than previously seen with polymer systems[7–10], and is comparable to the response time of piezoelectric devices.

X-ray crystallographic analysis of the crystals of **1** and **2** was carried out before and after ultraviolet-light irradiation (see Supplementary Methods and Supplementary Figs S5–S7), and supports the view that the crystalline photochromism is an unusual example of a single crystal to single crystal phototransformation[18–20] in terms of reversibility and macroscopic shape deformation. In conventional single crystal to single crystal transformations, changes in the cell dimensions are very small and macroscopic shape changes in the respective crystals are accordingly imperceptible. In contrast, the crystals studied here exhibit large and reversible macroscopic crystal-shape changes that accompany the photoisomerization of the molecules in the crystal lattice of the open-ring isomers. Even if closed-ring isomers are produced in the crystal lattice of the open-ring isomers, the space group remains the same (see Supplementary Tables S1 and S2). The specifics of the structural changes of the molecules in the crystals and the effect of these changes on the molecular packing in the crystals are therefore considered to have an important role in the large shape-changes observed.

Figure 4 shows the molecular packing of crystal **1** before ultraviolet irradiation, as seen when viewed from the (100) and (010) faces[15]. The change from a square to a lozenge shape indicates that the crystal contracts along the *c* axis and expands along the *b* axis. The molecular structures of the open- and closed-ring isomers of molecule **1** within the crystal, as determined by X-ray crystallographic analysis, indicate that the twisted thiophene rings become coplanar and that the thickness of each molecule is reduced as the molecule converts from the open- to the closed-ring isomer (see Supplementary Fig. S5). The co-facial packing of the thin layers of the planar closed-ring isomers along the *c* axis allows the molecules to be stacked one-by-one, resulting in the contraction along the *c* axis. A similar situation was observed for a crystal of **2** (see Supplementary Figs S6 and S7). That is, the transformation of each individual diarylethene molecule on photoirradiation influences the intermolecular interactions within the crystal so as to induce stacking of molecules, which then results in the macroscopic deformation of the crystal (see Supplementary Fig. S8).

Received 4 October 2006; accepted 10 February 2007.

1. Otsuka, K. & Wayman, C. M. *Shape Memory Materials* (Cambridge Univ. Press, Cambridge, UK, 1998).
2. Merian, E. Steric factors influencing the dyeing of hydrophobic fibers. *Text. Res. J.* **36**, 612–618 (1966).

3. Eisenbach, C. D. Isomerization of aromatic azo chromophores in poly(ethyl acrylate) networks and photomechanical effect. *Polymer* **21,** 1175–1179 (1980).
4. Agolini, F. & Gay, F. P. Synthesis and properties of azoaromatic polymers. *Macromolecules* **3,** 349–351 (1970).
5. Matějka, L., Ilavský, M., Dušek, K. & Wichterle, O. Photomechanical effects in crosslinked photochromic polymers. *Polymer* **22,** 1511–1515 (1981).
6. Irie, M. Photoresponsive polymers. *Adv. Polym. Sci* **94,** 28–67 (1990).
7. Finkelmann, H., Nishikawa, E., Pereira, G. G. & Warner, M. A new opto-mechanical effect in solids. *Phys. Rev. Lett.* **87,** 015501 (2001).
8. Yu, Y., Nakano, M. & Ikeda, T. Directed bending of a polymer film by light. *Nature* **425,** 145 (2003).
9. Ikeda, T., Nakano, M., Yu, Y., Tsutsumi, O. & Kanazawa, A. Anisotropic bending and unbending behavior of azobenzene liquid-crystalline gels by light exposure. *Adv. Mater.* **15,** 201–205 (2003).
10. Yu, Y., Nakano, M., Shishido, A., Shiono, T. & Ikeda, T. Effect of cross-linking density on photoinduced bending behavior of oriented liquid-crystalline network films containing azobenzene. *Chem. Mater.* **16,** 1637–1643 (2004).
11. Irie, M. Diarylethenes for memories and switches. *Chem. Rev.* **100,** 1685–1716 (2000).
12. Kobatake, S. & Irie, M. Single-crystalline photochromism of diarylethenes. *Bull. Chem. Soc. Jpn* **77,** 195–210 (2004).
13. Morimoto, M. & Irie, M. Photochromism of diarylethene single crystals: crystal structures and photochromic performance. *Chem. Commun.* 3895–3905 (2005).
14. Irie, M., Kobatake, S. & Horichi, M. Reversible surface morphology changes of a photochromic diarylethene single crystal by photoirradiation. *Science* **291,** 1769–1772 (2001).
15. Kobatake, S., Shibata, K., Uchida, K. & Irie, M. Photochromism of 1,2-bis(2-ethyl-5-phenyl-3-thienyl)perfluorocyclopentene in a single-crystalline phase. Conrotatory thermal cycloreversion of the closed-ring isomer. *J. Am. Chem. Soc.* **122,** 12135–12141 (2000).
16. Kobatake, S. *et al.* Absorption spectra of colored isomer of diarylethene in single crystals. *Chem. Lett. (Jpn)* **31,** 1224–1225 (2002).
17. Miyasaka, H., Nobuto, T., Itaya, A., Tamai, N. & Irie, M. Picosecond laser photolysis studies on photochromic dithienylethenes in solution and in crystalline phases. *Chem. Phys. Lett.* **269,** 281–285 (1997).
18. Nakanishi, H., Jones, W., Thomas, J. M., Hursthouse, M. B. & Motevalll, M. Static and dynamic single-crystal X-ray diffraction studies of some solid-state photodimerization reactions. *J. Phys. Chem.* **85,** 3636–3642 (1981).
19. Novak, K., Enkelmann, V., Wegner, G. & Wagener, K. B. Crystallographic study of single crystal to single crystal photodimerization and its thermal reverse reaction. *Angew. Chem. Int. Edn Engl.* **32,** 1614–1616 (1993).
20. Toh, N. L., Nagarathinam, M. & Vittal, J. J. Topochemical photodimerization in the coordination polymer $[\{(CF_3CO_2)(\mu\text{-}O_2CCH_3)Zn\}_2(\mu\text{-bpe})_2]_n$ through single-crystal to single-crystal transformation. *Angew. Chem. Int. Edn* **44,** 2237–2241 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The present work was supported by Grant-in-Aid for Scientific Research on Priority Areas and Nanotechnology Support Project from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

Author Contributions M.I. designed the study, and wrote the paper. S.K. and T.I. performed experiments on compound **1**. S.T. and H.M. performed experiments on compound **2**.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.I. (irie@cstf.kyushu-u.ac.jp).

LETTERS

Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems

Gregory S. Engel[1,2], Tessa R. Calhoun[1,2], Elizabeth L. Read[1,2], Tae-Kyu Ahn[1,2], Tomáš Mančal[1,2]†, Yuan-Chung Cheng[1,2], Robert E. Blankenship[3,4] & Graham R. Fleming[1,2]

Photosynthetic complexes are exquisitely tuned to capture solar light efficiently, and then transmit the excitation energy to reaction centres, where long term energy storage is initiated. The energy transfer mechanism is often described by semiclassical models that invoke 'hopping' of excited-state populations along discrete energy levels[1,2]. Two-dimensional Fourier transform electronic spectroscopy[3–5] has mapped[6] these energy levels and their coupling in the Fenna–Matthews–Olson (FMO) bacteriochlorophyll complex, which is found in green sulphur bacteria and acts as an energy 'wire' connecting a large peripheral light-harvesting antenna, the chlorosome, to the reaction centre[7–9]. The spectroscopic data clearly document the dependence of the dominant energy transport pathways on the spatial properties of the excited-state wavefunctions of the whole bacteriochlorophyll complex[6,10]. But the intricate dynamics of quantum coherence, which has no classical analogue, was largely neglected in the analyses—even though electronic energy transfer involving oscillatory populations of donors and acceptors was first discussed more than 70 years ago[11], and electronic quantum beats arising from quantum coherence in photosynthetic complexes have been predicted[12,13] and indirectly observed[14]. Here we extend previous two-dimensional electronic spectroscopy investigations of the FMO bacteriochlorophyll complex, and obtain direct evidence for remarkably long-lived electronic quantum coherence playing an important part in energy transfer processes within this system. The quantum coherence manifests itself in characteristic, directly observable quantum beating signals among the excitons within the *Chlorobium tepidum* FMO complex at 77 K. This wavelike characteristic of the energy transfer within the photosynthetic complex can explain its extreme efficiency, in that it allows the complexes to sample vast areas of phase space to find the most efficient path.

In two-dimensional Fourier transform electronic spectroscopy, three pulses and a strongly attenuated local oscillator are incident on the sample. (For a detailed description of the method and the apparatus used in this study, see refs 3, 5, 15.) In the simplest time domain representation, the first pulse creates a coherence that evolves for time τ, then the second pulse creates an excited-state population that evolves for time T, and the third pulse creates a coherence that accumulates phase in the opposite direction for time t before rephasing occurs and a signal pulse is emitted in the unique phase-matched direction. The full electric field of the signal pulse is measured through heterodyne detection using spectral interferometry. In a frequency domain representation, two-dimensional Fourier transform electronic spectroscopy probes electronic couplings and energy transfer in molecules by mapping how excitations before the population time T affect emission after the population time. The coherence wavelength represents the initial excitation, while the rephasing wavelength can be thought of as the subsequent emission. Without coupling, contributions from excited-state absorption and emission cancel each other, yielding no off-diagonal peaks in the spectrum that signal such coupling. But in the presence of coupling, the cancellation is no longer complete and a so-called cross-peak emerges[16]. Two-dimensional spectroscopy thus provides an excellent probe of the coupling between energy levels.

In the present experiment, we use two-dimensional electronic spectroscopy to observe oscillations caused by electronic coherence evolving during the population time in FMO. Such quantum coherence, a coherent superposition of electronic states analogous to a nuclear wavepacket in the vibrational regime, is formed when the system is initially excited by a short light pulse with a spectrum that spans multiple exciton transitions. Theoretical predictions indicate that both the amplitudes and shapes of peaks will contain beating signals with frequencies corresponding to the differences in energy between component exciton states[17].

To observe the quantum beats, two-dimensional spectra were taken at 33 population times T, ranging from 0 to 660 fs. Representative spectra are shown in Fig. 1 and a video of the spectral evolution is included in the Supplementary Information. In these spectra, the lowest-energy exciton gives rise to a diagonal peak near 825 nm that clearly oscillates: its amplitude grows, fades, and subsequently grows again. The peak's shape evolves with these oscillations, becoming more elongated when weaker and rounder when the signal amplitude intensifies. The associated cross-peak amplitude also appears to oscillate. Surprisingly, the quantum beating lasts for 660 fs. This observation contrasts with the general assumption that the coherences responsible for such oscillations are destroyed very rapidly, and that population relaxation proceeds with complete destruction of coherence[2] (so that the transfer of electronic coherence between excitons during relaxation is usually ignored[2,18,19]).

We note that this issue has been discussed theoretically, and that some theoretical models include coherence among both donors and acceptors[20]; however, coherence must be treated between all chromophores to ensure that the model will accurately reproduce the dynamics of the system[21]. In the case of FMO, theoretical models indicate that electronic coherence should dephase on the timescale of the initial population transfer, established by experiment to be less than 250 fs for all but excitons 1 and 3 (refs 2, 22). The strong quantum beating that we observe to last for at least 660 fs clearly exceeds the model predictions. We believe that to account for this long-lived coherence and provide an accurate description of the system, the protein must have a more active role in a realistic bath model; that is, it must be allowed to interact with both donors and

[1]Department of Chemistry & QB3 Institute, University of California, Berkeley, [2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [3]Department of Biology, [4]Department of Chemistry, Washington University, St Louis, Missouri 63130, USA. †Present address: Institute of Physics of Charles University, 12116 Prague 2, Czech Republic.

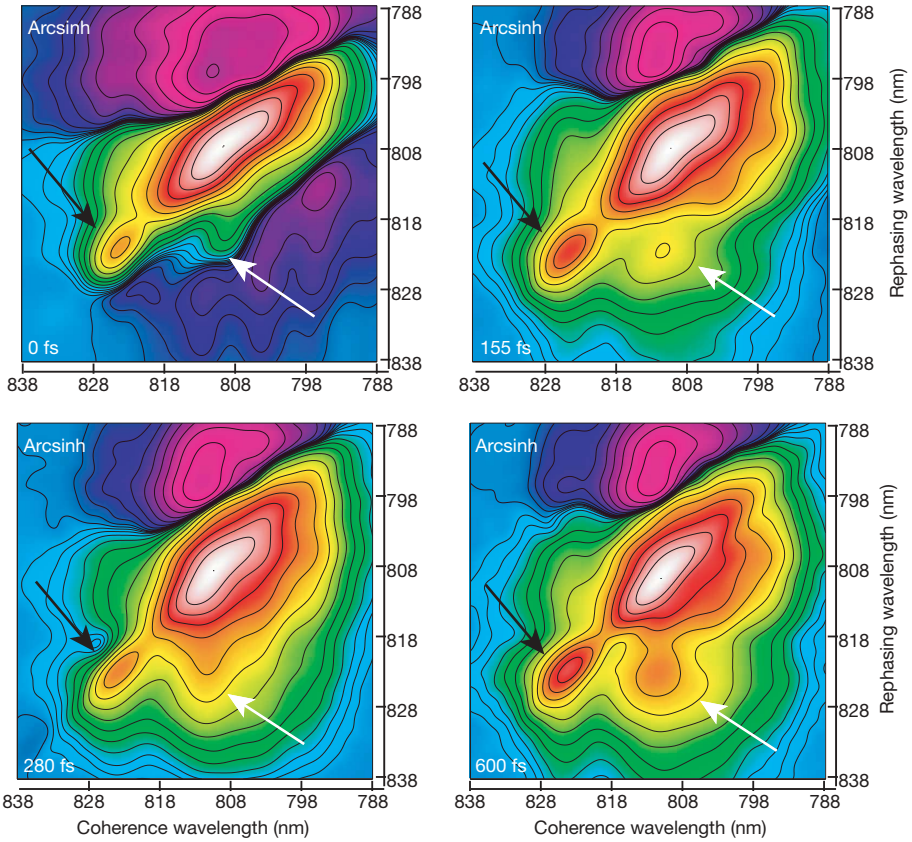


Figure 1 | Two-dimensional electronic spectra of FMO. Selected two-dimensional electronic spectra of FMO are shown at population times from $T = 0$ to 600 fs demonstrating the emergence of the exciton 1–3 cross-peak (white arrows), amplitude oscillation of the exciton 1 diagonal peak (black arrows), the change in lowest-energy exciton peak shape and the oscillation of the 1–3 cross-peak amplitude. The data are shown with an arcsinh coloration to highlight smaller features: amplitude increases from blue to white (for a three-dimensional representation of the coloration see Fig. 3a).

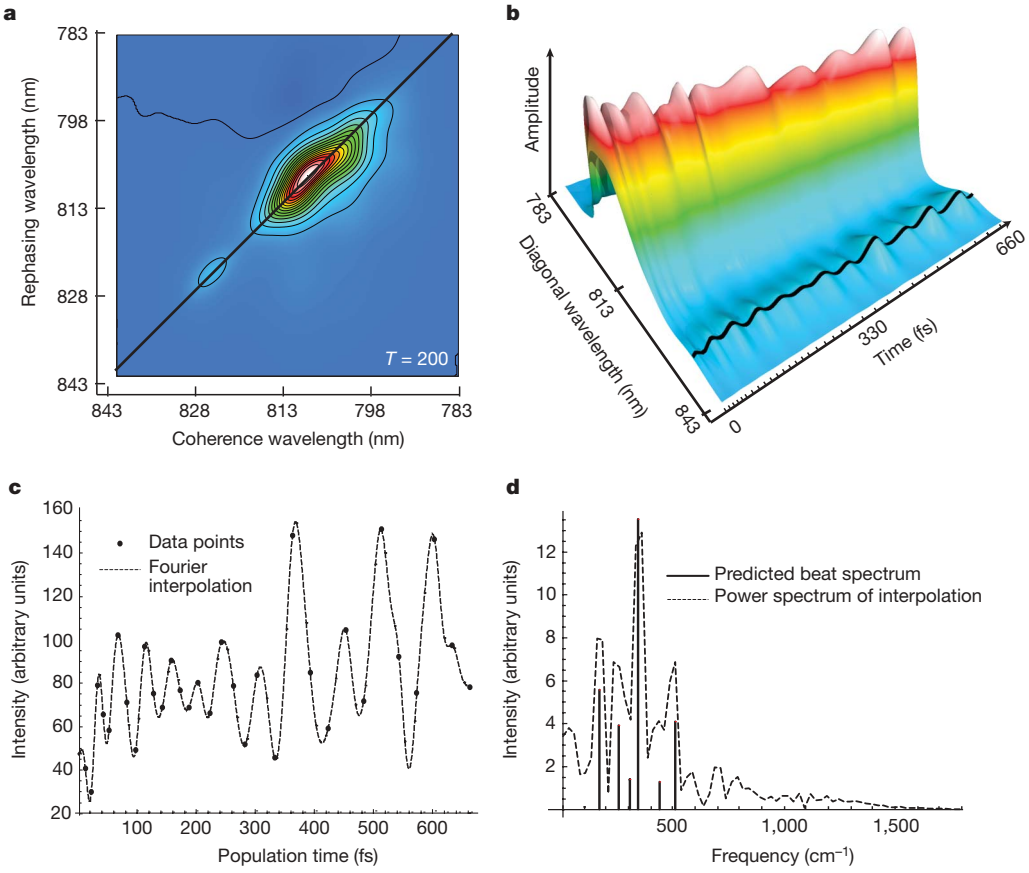


Figure 2 | Electronic coherence beating. a, A representative two-dimensional electronic spectrum with a line across the main diagonal peak. The amplitude along this diagonal line is plotted against population time in **b** with a black line covering the exciton 1 peak amplitude; the data are scaled by a smooth function effectively normalizing the data without affecting oscillations. A spline interpolation is used to connect the spectra; the times at which spectra were taken are denoted by tick marks along the time axis. **c**, The amplitude of the peak corresponding to exciton 1 shown with a dotted Fourier interpolation. **d**, The power spectrum of the Fourier interpolation in **c** is plotted with the theoretical spectrum showing beats between exciton 1 and excitons 2–7.

acceptors, to enable coherence transfer and possibly the generation of new coherences.

Our further investigation of these spectral oscillations focuses on the best-resolved spectral features in the two-dimensional electronic spectra, that is, the lowest exciton diagonal peak and associated cross-peak. In Fig. 2, we show the amplitude oscillations with time along the main diagonal of the spectrum. Because of the non-uniform spacing of the data points, converting the data to a unique power spectrum is not possible; instead, we map the data onto a Fourier subspace using a non-uniform fast Fourier transform algorithm[23]. The sampled amplitude variations of the lowest-energy exciton are shown with a Fourier interpolation from the subspace that was selected to be maximally consistent with the excitonic model. The theoretical excitonic coherence line spectrum was calculated using the exciton energies from the hamiltonian in ref. 6, and the relative amplitudes were calculated using orientationally averaged response magnitudes for the associated cross-peak pathways[16,24,25].

The agreement between the data, the Fourier interpolation and the theoretical exciton spectrum demonstrates that the quantum beating observed is fully consistent with electronic coherence (analyses of beating in more peaks are shown in the Supplementary Information), and we present comparisons to an independent theoretical lineshape prediction to confirm this conclusion. Further, because the predicted orientational factor is identical for a diagonal exciton beating and for a cross-peak, that the amplitude of the beat is of the order of the amplitude of the cross-peak even at long times (>500 fs) indicates that the electronic coherence can play a significant role in determining the overall relaxation dynamics within the protein complex. This observation suggests that coherence relaxation pathways, including coherence transfer, should no longer be disregarded in theoretical models of photosynthetic protein complexes.

A predicted[17] signature of quantum beating in a dimer system is out-of-phase modulation of the diagonal and antidiagonal widths of a peak relative to its amplitude, with the peak predicted to become rounder as it gets stronger. Figure 3 documents precisely this behaviour in the lowest exciton state. We note that if this oscillation were due to vibrational wavepacket motion, the exciton peak would be expected instead to oscillate in frequency but maintain constant volume. The peak shape oscillations we see have not previously been observed in two-dimensional spectra, in which the peak width along the antidiagonal direction is generally an indicator of the extent of homogeneous broadening, whereas the peak width along the diagonal provides a measure of inhomogeneous broadening. In this context, the width modulation we observe would be akin to losing memory of the initial state, and subsequently regaining that memory. Clearly, this interpretation is not applicable while coherence persists. The unusual nature of the width modulation evident in our data and its agreement with the predicted[17] characteristics of quantum beating further strengthens our conclusion that the beating we measure is due to excitonic quantum coherence.

Figure 4 shows that a Fourier interpolation of the beating spectrum apparent in the cross-peak between excitons 1 and 3 can again be explained by the expected exciton beating spectrum. The beating signal shows frequency components from all excitons coupled to either excitons 1 or 3, and a very strong component from the beat frequency between the two coupled excitons (1 and 3) that give rise to the cross peak. However, the cross-peak amplitude and associated beating do not appear strongly at time $T = 0$; this provides strong evidence for the existence of electronic coherence transfer in this system, with one superposition, for example between excitons 3 and 5, transferred into a superposition between excitons 3 and 1.

The observations illustrated in Figs 2–4 clearly demonstrate that any full description of FMO dynamics needs to account for coherence between donors and acceptors, and for coherence transfer on a similar timescale to population relaxation. While our spectra are recorded at 77 K and the coherence among members of the ensemble will dephase faster at higher temperatures, quantum coherence will also be important at higher temperatures. This is because the underlying hamiltonian, dynamics and relaxation pathways associated with superposition states of individual complexes do not depend on dephasing of the ensemble properties observed in a measurement. In an ensemble sense, the presence of quantum coherence means that the mean square displacement of the initial excitation from a quantum walk increases quadratically with time rather than linearly as in diffusive hopping of excitation[26].

But a more microscopic picture of the implications of such quantum coherence for energy transfer can be found by considering an individual complex: superposition states formed during a fast excitation event allow the excitation to reversibly sample relaxation rates from all component exciton states, thereby efficiently directing the energy transfer to find the most effective sink for the excitation energy (which, in the isolated FMO complex, is the lowest energy state). When viewed in this way, the system is essentially performing a single quantum computation, sensing many states simultaneously and selecting the correct answer, as indicated by the efficiency of the energy transfer. In the presence of quantum coherence transfer, such an operation is analogous to Grover's algorithm, with the hamiltonian describing both relaxation to the lowest energy state and coherence transfer (refilling the coherence lost from the transfer to the lowest-energy state)[27]; such a scheme can provide efficiency beyond

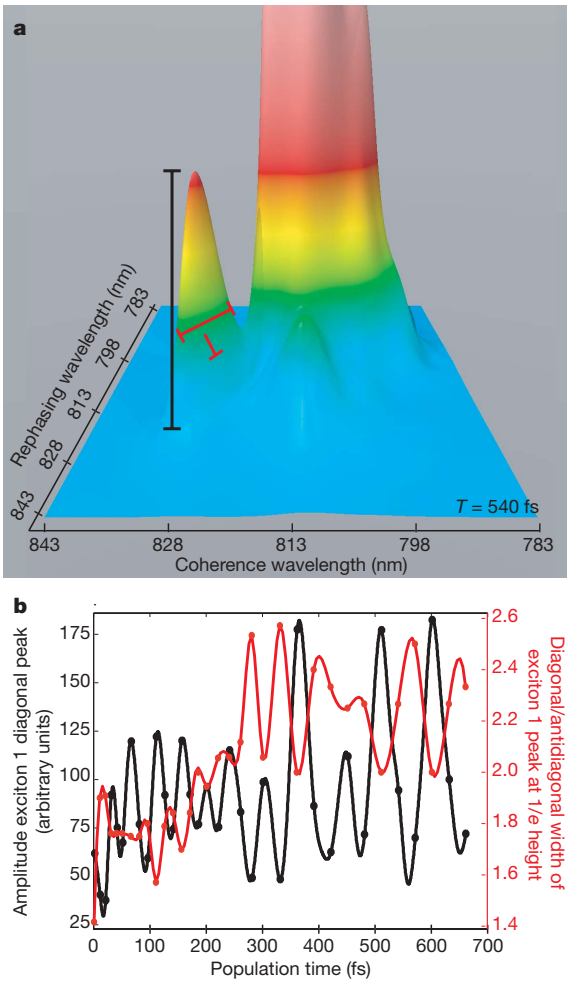


Figure 3 | Characteristic anticorrelation between peak amplitude and width. The anticorrelation shown in **b** between the amplitude of the diagonal exciton peak (black line in **a**) and the ratio of the diagonal to anti-diagonal widths of the peak (red lines in **a**) is a characteristic predicted from theory for exciton quantum beating[17]. This pattern would not arise from phonon coupling and highlights the change in integrated line strength associated with quantum beating.

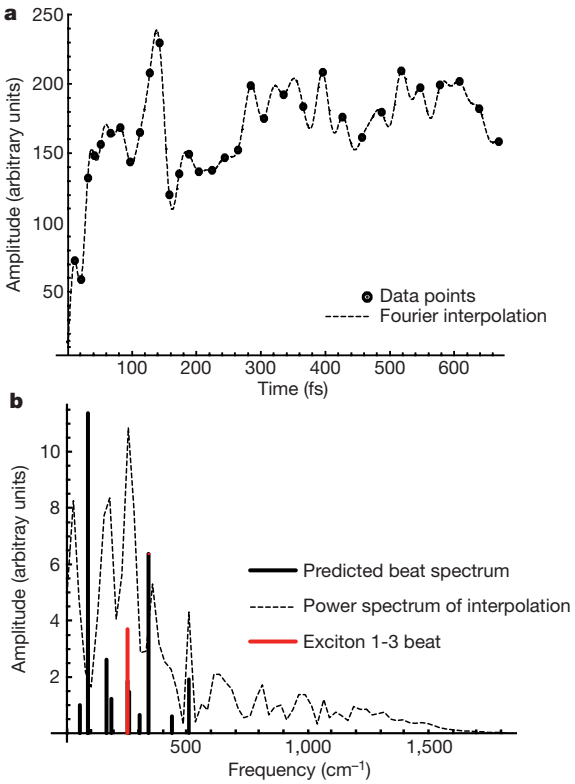


Figure 4 | Quantum beating in cross peaks. a, The raw amplitude of the exciton 1–3 cross-peak, with a Fourier interpolation of the points (dotted line). **b**, The power spectrum of this interpolation (dotted line), the exciton beating line spectra of both excitons 1 and 3 (black), and the 1–3 beat frequency (red). We expect that the other frequencies may couple to this cross-peak but that the dominant frequency corresponds to the red transition. The apparent low-frequency peak is due to the growth of the cross-peak amplitude and appears as a peak because the data were de-meaned (mean subtracted from the data) before the transform to improve numerical accuracy, pinning the zero-frequency component to zero.

that of a classical search algorithm. This mechanism contrasts with a semiclassical 'hopping' mechanism through which the excitation moves stepwise from exciton state to exciton state, dissipating energy at each step, which would be similar to a classical search where only one state can be occupied at any one time. Such a mechanism also raises the possibility of non-local events, although more detailed analysis is needed before we can determine whether such effects are present in FMO.

The FMO light-harvesting complex provides an opportunity to apply more complete energy transfer theories that invoke non-markovian dynamics and include coherence transfer. Such theories need to include wavelike energy motion owing to long-lived coherence terms, alongside the population transfer included in current models. Further, the observed preservation of coherence in this photosynthetic system requires us to redefine our description of the role of electron–phonon interactions within photosynthetic proteins. In particular, the protein may not only enforce the structure that gives rise to the couplings, but also modulate those couplings with motions of charged residues and changing local dielectric environments, which will change exciton energies and promote coherence transfer.

METHODS

Sample preparation. The FMO sample was isolated from *Chlorbium tepidum* as published previously[9]. The sample was dissolved in a buffer of 800 mM tris/HCl pH 8.0, 50 mM NaCl with 0.1% lauryldimethylamine oxide as a detergent. The sample was then mixed 65:35 v/v in glycerol, placed in a 200 μm quartz cell (Starna). The sample was cooled in a cryostat (Oxford Instruments) to 77 K. The absorbance of the sample at 805 nm was then measured to be 0.16.

Data acquisition. A home-built oscillator was used to seed a home-built regenerative amplifier to produce a 3.4 kHz pulse train of 41 fs pulses centred about 808 nm with a spectral width of 31 nm full-width at half-maximum (FWHM)[28]. The stability of the laser system through the data acquisition period was measured to be 0.28% to 0.44%. The laser pulse width was measured with both autocorrelation (38 fs FWHM) and frequency resolved optical gating (FROG) (41 fs FWHM).

A diffractive optic was used to create two pairs of phase-locked beams, and all beams were incident on the same optics, aside from the one-degree fused silica wedges (Almaz Optics) used for delay stages. The delay stages were calibrated using spectral interferometry and have been shown to have attosecond stability and reproducibility[3,29]. During data collection, the coherence time was stepped in 4 fs steps from −600 to 600 fs. Population times were sampled at 0, 10, 20, 30, 40, 50, 65, 80, 95, 110, 125, 140, 155, 170, 185, 200, 220, 240, 260, 280, 300, 330, 360, 390, 420, 450, 480, 510, 540, 570, 600, 630 and 660 fs. Additional points were taken at −1,000 fs to ensure that the signal observed was not an artefact. The sample was moved after every third data point, and some data points were repeated to ensure that the sample was not degrading during measurements.

The local oscillator was attenuated with a neutral density filter combination of absorbance 3.5 at 800 nm. The total power incident on the sample was 15 nJ (5 nJ per pulse) focused to a spot size of less than 70 μm. The resulting signal and heterodyne beam were frequency-resolved with a spectrometer (Acton Research) and captured on a 1,340 × 5 pixel region of our charge-coupled device (CCD) array (Princeton Instruments). Scatter subtraction, Fourier windowing and transformation to frequency–frequency space was done as reported previously[3,15]. Pump–probe data were taken for accurate phasing of the two dimensional spectra.

Data modelling. To model the beat patterns in the data series, a non-uniform fast Fourier transform algorithm was used to search for Fourier power series that matched the data and were consistent with expected power spectrum. Because of the unequal spacing, a unique answer does not exist. The non-uniform fast Fourier transform power spectra were estimated using a modified 4th-Jackson kernel, and the results were then transformed back using a standard fast Fourier transform library (FFTW; www.fftw.org)[23,30]. To scale the data and eliminate spectral components from population dynamics, a smooth function with an exponential growth of the order of the pulse width and an exponential decay matched to the population dynamics in the main peak was used.

Received 13 October 2006; accepted 14 February 2007.

1. Blankenship, R. E. *Molecular Mechanisms of Photosynthesis* (Blackwell Science, Oxford/Malden, 2002).
2. van Amerongen, H., Valkunas, L. & van Grondelle, R. *Photosynthetic Excitons* (World Scientific, Singapore/River Edge, 2000).
3. Brixner, T., Mančal, T., Stiopkin, I. V. & Fleming, G. R. Phase-stabilized two-dimensional electronic spectroscopy. *J. Chem. Phys.* **121,** 4221–4236 (2004).
4. Jonas, D. M. Two-dimensional femtosecond spectroscopy. *Annu. Rev. Phys. Chem.* **54,** 425–463 (2003).
5. Cowan, M. L., Ogilvie, J. P. & Miller, R. J. D. Two-dimensional spectroscopy using diffractive optics based phased-locked photon echoes. *Chem. Phys. Lett.* **386,** 184–189 (2004).
6. Brixner, T. *et al.* Two-dimensional spectroscopy of electronic couplings in photosynthesis. *Nature* **434,** 625–628 (2005).
7. Fenna, R. E. & Matthews, B. W. Chlorophyll arrangement in a bacteriochlorophyll protein from *Chlorobium limicola*. *Nature* **258,** 573–577 (1975).
8. Li, Y. F., Zhou, W. L., Blankenship, R. E. & Allen, J. P. Crystal structure of the bacteriochlorophyll a protein from *Chlorobium tepidum*. *J. Mol. Biol.* **271,** 456–471 (1997).
9. Camara-Artigas, A., Blankenship, R. E. & Allen, J. P. The structure of the FMO protein from *Chlorobium tepidum* at 2.2 angstrom resolution. *Photosynth. Res.* **75,** 49–55 (2003).
10. Cho, M. H. *et al.* Exciton analysis in 2D electronic spectroscopy. *J. Phys. Chem. B* **109,** 10542–10556 (2005).
11. Perrin, F. Thoérie quantique des transferts d'activation entre molécules de même espèce. Cas des solutions fluorescentes. *Ann. Phys. (Paris)* **17,** 283–314 (1932).
12. Knox, R. S. Electronic excitation transfer in the photosynthetic unit: Reflections on work of William Arnold. *Photosynth. Res.* **48,** 35–39 (1996).
13. Leegwater, J. A. Coherent versus incoherent energy transfer and trapping in photosynthetic antenna complexes. *J. Phys. Chem.* **100,** 14403–14409 (1996).
14. Savikhin, S., Buck, D. R. & Struve, W. S. Oscillating anisotropies in a bacteriochlorophyll protein: Evidence for quantum beating between exciton levels. *Chem. Phys.* **223,** 303–312 (1997).
15. Brixner, T., Stiopkin, I. V. & Fleming, G. R. Tunable two-dimensional femtosecond spectroscopy. *Opt. Lett.* **29,** 884–886 (2004).
16. Cho, M. H. & Fleming, G. R. The integrated photon echo and solvation dynamics. II. Peak shifts and two-dimensional photon echo of a coupled chromophore system. *J. Chem. Phys.* **123,** 114506 (2005).
17. Pisliakov, A. V., Mančal, T. & Fleming, G. R. Two-dimensional optical three-pulse photon echo spectroscopy. II. Signatures of coherent electronic motion and

exciton population transfer in dimer two-dimensional spectra. *J. Chem. Phys.* **124,** 234505 (2006).
18. Abramavicius, D., Valkunas, L. & van Grondelle, R. Exciton dynamics in ring-like photosynthetic light-harvesting complexes: A hopping model. *Phys. Chem. Chem. Phys.* **6,** 3097–3105 (2004).
19. Renger, T., May, V. & Kuhn, O. Ultrafast excitation energy transfer dynamics in photosynthetic pigment-protein complexes. *Phys. Rep. Rev. Phys. Lett.* **343,** 138–254 (2001).
20. Jang, S. J., Newton, M. D. & Silbey, R. J. Multichromophoric Forster resonance energy transfer. *Phys. Rev. Lett.* **92,** 9312–9323 (2004).
21. Novoderezhkin, V., Wendling, M. & van Grondelle, R. Intra- and interband transfers in the b800-b850 antenna of *Rhodospirillum molischianum*: Redfield theory modeling of polarized pump-probe kinetics. *J. Phys. Chem. B* **107,** 11534–11548 (2003).
22. Vulto, S. I. E. *et al.* Excited state dynamics in FMO antenna complexes from photosynthetic green sulfur bacteria: A kinetic model. *J. Phys. Chem. B* **103,** 8153–8161 (1999).
23. Potts, D. & Kunis, S. Stability results for scattered data interpolation by trigonometric polynomials. ⟨http://arxiv.org/pdf/math.NA/0702019⟩ (2007).
24. Dreyer, J., Moran, A. M. & Mukamel, S. Tensor components in three pulse vibrational echoes of a rigid dipeptide. *Bull. Korean Chem. Soc.* **24,** 1091–1096 (2003).
25. Hochstrasser, R. M. Two-dimensional IR-spectroscopy: polarization anisotropy effects. *Chem. Phys.* **266,** 273–284 (2001).
26. Kempe, J. Quantum random walks: An introductory overview. *Contemp. Phys.* **44,** 307–327 (2003).
27. Grover, L. K. Quantum mechanics helps in searching for a needle in a haystack. *Phys. Rev. Lett.* **79,** 325–328 (1997).
28. Joo, T., Jia, Y. W. & Fleming, G. R. Ti-sapphire regenerative amplifier for ultrashort high-power multikilohertz pulses without an external stretcher. *Opt. Lett.* **20,** 389–391 (1995).
29. Lepetit, L., Cheriaux, G. & Joffre, M. Linear techniques of phase measurement by femtosecond spectral interferometry for applications in spectroscopy. *J. Opt. Soc. Am. B* **12,** 2467–2474 (1995).
30. Frigo, M. & Johnson, S. G. The design and implementation of fftw3. *Proc. IEEE* **93,** 216–231 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Zigmantas for discussions and J. Wen for purification of the sample. This work was supported by the DOE (at LBNL, UC Berkeley and Washington Univ.). G.S.E. thanks the Miller Institute for Basic Research in Science for support. T.-K.A. was supported by the Korea Research Foundation Grant funded by the Korean government (MOEHRD).

Author Contributions G.S.E, T.R.C., T.-K.A. and E.L.R. prepared the cryogenic sample and collected the data; G.S.E., E.L.R, T.M. and Y.-C.C. performed the data analysis. R.E.B. grew, isolated and purified the FMO sample. G.S.E. wrote the paper, and all authors discussed the results and commented on the manuscript. G.R.F. provided guidance throughout the experiment and analysis and helped to write the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.R.F. (grfleming@lbl.gov).

LETTERS

A periodic shear-heating mechanism for intermediate-depth earthquakes in the mantle

Peter B. Kelemen[1] & Greg Hirth[2]

Intermediate-depth earthquakes[1], at depths of 50–300 km in subduction zones, occur below the brittle–ductile transition, where high pressures render frictional failure unlikely. Their location approximately coincides with 600 to 800 °C isotherms in thermal models[2], suggesting a thermally activated mechanism for their origin. Some earthquakes may occur by frictional failure owing to high pore pressure that might result from metamorphic dehydration[2–5]. Because some intermediate-depth earthquakes occur ~30 to 50 km below the palaeo-sea floor[6], however, the hydrous minerals required for the dehydration mechanism may not be present. Here we present an alternative mechanism to explain such earthquakes, involving the onset of highly localized viscous creep in pre-existing, fine-grained shear zones. Our numerical model uses olivine flow laws for a fine-grained, viscous shear zone in a coarse-grained, elastic half space, with initial temperatures from 600–800 °C and background strain rates of 10^{-12} to 10^{-15} s^{-1}. When shear heating becomes important, strain rate and temperature increase rapidly to over 1 s^{-1} and 1,400 °C. The stress then drops dramatically, followed by low strain rates and cooling. Continued far-field deformation produces a quasi-periodic series of such instabilities.

There is geological and geophysical evidence for our proposed mechanism. In shallow mantle peridotite there are mylonitic shear zones (1–100 μm olivine) within residual peridotite (~5–10 mm olivine)[7–9]. Fine-grained shear zones form by recrystallization and annealing of gouge along oceanic transforms, subduction-related thrusts, and faults at the 'outer rise'. Outer-rise earthquakes extend to ≥40 km[10], and their reactivation may play a role in intermediate-depth earthquakes[11]. Mixing of olivine and pyroxene in mylonites[12] by viscous or cataclastic processes limits grain growth, owing to grain size pinning. Thus, fine-grained shear zones probably persist for 10^8 to 10^9 years at shallow mantle temperatures[13]. Fine-grained shear zones have a significantly lower viscosity than surrounding rock when heated to higher temperatures. Some peridotite shear zones reach extremely high temperatures, sufficient to form veins of melt that rapidly cool and are chilled to glass, known as 'pseudotachylites' (see ref. 14), and some deep earthquakes may involve melting[15].

Shear-heating instabilities have been the subject of previous studies[16–21]. Positive feedback between strain-rate-dependent shear heating and temperature-dependent strain rate drives rapidly increasing strain rate. These studies indicate that instability can occur, but have not yielded a model for periodic instability. Recent work has investigated unstable and stable deformation involving shear heating and grain-size-sensitive creep[22–24] and localization with transient heating arising from low-temperature plasticity[25].

Inspired by Whitehead and Gans[16], we model periodic shear heating in pre-existing fine-grained shear zones. Our work differs from previous studies in (1) positing pre-existing, fine-grained shear zones which localize grain-size-sensitive creep, rather than *ad hoc*, localized temperature perturbations, (2) using experimental flow laws, including grain boundary sliding, for olivine, and (3) limiting the maximum temperature.

Our model tracks the behaviour of a fine-grained, viscous shear zone with width w_{sz}, adjacent to a block of elastic peridotite with width w_{ez}, and a constant velocity on the far side of the elastic zone (Fig. 1). To model strain rate in the shear zone, we use olivine flow laws for dislocation creep, diffusion creep, and grain boundary sliding[26], plus low-temperature plasticity[27], and sum these to determine the total viscous strain rate, $\dot{\varepsilon}_{viscous}$. The Supplementary Information reports the flow laws.

Shear heating and diffusion control the shear-zone temperature:

$$dT/dt = -\kappa[d^2T/dx^2] + \sigma\,\dot{\varepsilon}_{viscous}/(C_p\rho) \quad (1)$$

for $T < 1{,}400$ °C, where κ is thermal diffusivity, C_p is heat capacity and ρ is density. The Crank Nicholson method is used to calculate temperature over a 10-km-wide region around the shear zone.

At and above ~1,400 °C in shallow mantle peridotite, melting would rapidly decrease viscosity, stress, and the amount of shear heating, so we impose an upper bound for temperature by setting the second term in equation (1) to zero when the shear zone reaches 1,400 °C. As discussed in the Supplementary Information, a more complex formulation would incorporate the effect of melting on viscosity and temperature.

We used far-field strain rates of 10^{-12} to 10^{-15} s^{-1}, set by the ratio of the far-field velocity to the elastic zone width, w_{ez}. Total strain is

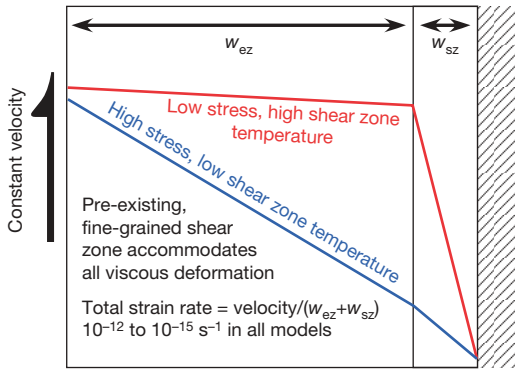


Figure 1 | Schematic illustration of some numerical model conditions and results. Blue lines show displacement at high stress and low temperature, before any shear heating instability. At this point, viscous deformation in the shear zone is negligible compared to the far-field displacement and deformation is distributed. Red lines show displacement at low stress and high temperature, after the first shear heating instability has occurred. At this time, viscous displacement has accommodated most of the far-field displacement, and deformation of the elastic zone has decreased.

[1]Department of Earth and Environmental Sciences, Columbia University, Lamont Doherty Earth Observatory, Palisades, New York 10964, USA. [2]Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA.

the product of total time elapsed and far-field velocity divided by $w_{ez} + w_{sz}$.

Stress is constant everywhere at each time step, and is the 'elastic stress':

$$\sigma = G\left[(w_{ez} + w_{sz})\cdot\varepsilon_{total} - w_{sz}\varepsilon_{viscous}\right]/(w_{ez} + w_{sz}) \quad (2)$$

where the shear modulus $G = 5 \times 10^{10}$ Pa (inertial terms are discussed in the Supplementary Information).

Our results are essentially independent of grain size in the shear zone for initial grain size $\ll$10 mm. Because grain growth may be limited by second-phase pinning[9,12], we could justifiably impose a maximum grain size of 10 μm in our models. However, we model grain-size evolution in the shear zone and in a fictive 'wall rock' because a crucial assumption is that viscous deformation remains localized within fine-grained shear zones. This assumption would be violated if (1) extensive viscous deformation took place via a grain-size-independent mechanism (dislocation creep, low-temperature plasticity), or (2) if shear-zone grain growth during high-temperature, low-stress periods (or wall-rock grain-size reduction during high-stress periods) eliminated the grain-size contrast between shear zone and 'wall rock'. The 'wall rock' undergoes the same stress and temperature history as the shear zone, but has an initial grain size of 10 mm. We used a steady-state grain-size piezometer for olivine, where the grain size is attained after a critical strain at a given stress, and an expression for diffusive grain growth (see Supplementary Information).

Typical results are illustrated in Figs 2 and 3, and in Supplementary Fig. S1. Shear heating events occur with a period of 200 to 250 years (Figs 2a, 3a). These are driven by accelerating viscous strain and shear heating (Figs 2b, 3b), which result from increasing stress between events (Fig. 2c) as viscous displacement lags behind far-field displacement (Figs 2d, 3d). About half the stress drop and displacement during shear heating events occurs while strain rates are greater than 1 s^{-1} (Figs 3c, d). Thus, our model includes pre- and post-seismic creep as well as seismic deformation. The grain size in the shear zone remains small despite grain growth during periods of high temperature and low stress (Fig. 2e). Displacement in the shear zone is much larger than in the fictive 'wall rock' (Figs 2d, f), supporting our assumption that viscous strain in the shallow mantle can remain localized within pre-existing, fine-grained shear zones.

Between heating events, thermal diffusion cools the shear zone and heats its surroundings. The shear zone never cools to its initial temperature; instead, the starting temperature increases for each successive heating event (Fig. 2a). Once the adjacent 'wall rock' temperature reaches ~850 °C, subsequent displacement is accommodated by steady-state creep in the shear zone, and instabilities cease.

To first order, the period between heating events is proportional to the reciprocal of the far-field strain rate (see Supplementary Fig. S2). In model runs with lower strain rates, more instabilities occur before the system reaches steady-state, owing to longer cooling intervals. Decreasing the initial temperature extends the time interval between the onset of shear heating events and the transition to steady, viscous deformation in the shear zone. At a given far-field strain rate, decreasing the elastic-zone width w_{ez} decreases the period between shear heating events and the viscous displacement during events, and also tends to stabilize the model system.

A steady-state approximation provides a first-order confirmation of the numerical results. At steady state, the diffusive temperature flux is equal to the shear heating flux, which can be approximated with a linear thermal gradient around the shear zone, extending over a diffusion-zone width w_{dz}. The grain size is set by a steady-state piezometer[28] (Supplementary equation S6), and the strain rate is constant. We used the grain boundary sliding mechanism (Supplementary equation S3), the dominant creep mechanism in our model. These assumptions yield:

$$\sigma = \left\{\dot{\varepsilon}/\left[4.4 \times 10^{-9} A e^{-Q/(RT_{ss})}\right]\right\}^{1/6.1} \quad (3)$$

This expression is nonlinear because steady-state temperature (T_{ss}) and strain rate are functions of stress (Supplementary equations S3 and S12).

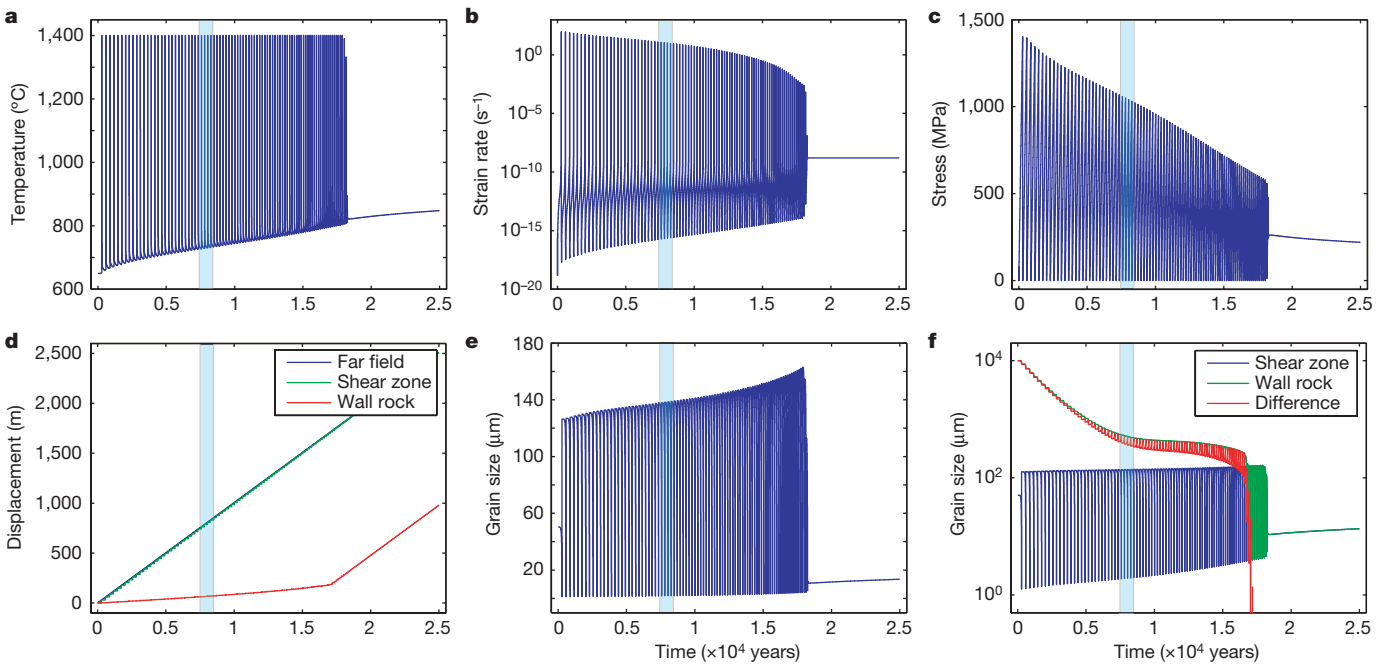


Figure 2 | Numerical model results for $T_0 = 650$ °C, $w_{ez} = 1$ km, $w_{sz} = 2$ m, $\sigma_0 = 80$ MPa, $gs_0 \approx 50$ μm, adjacent 'wall rock' grain size of 10 mm, and far-field velocity of 0.1 m yr^{-1}. The light blue rectangle highlights the time interval from 7,600 to 8,600 years, shown in detail in Fig. 3 and Supplementary Fig. S1. **a**, Time versus temperature. **b**, Time versus strain rate. **c**, Time versus stress. **d**, Time versus displacement. 'Far field', displacement on the free edge of the elastic zone (left side of Fig. 1); 'shear zone', displacement on the moving side of the shear zone (cumulative shear-zone strain × w_{sz} × time; left side of shear zone in Fig. 1); 'wall rock', displacement in a fictive 'wall rock' on the left side of the shear zone with an initial grain size of 10 mm that undergoes the same temperature and stress history as the shear zone. **e**, Time versus shear-zone grain size. **f**, Time versus shear-zone and wall-rock grain size, with the difference ('wall rock' minus grain size) shown in red.

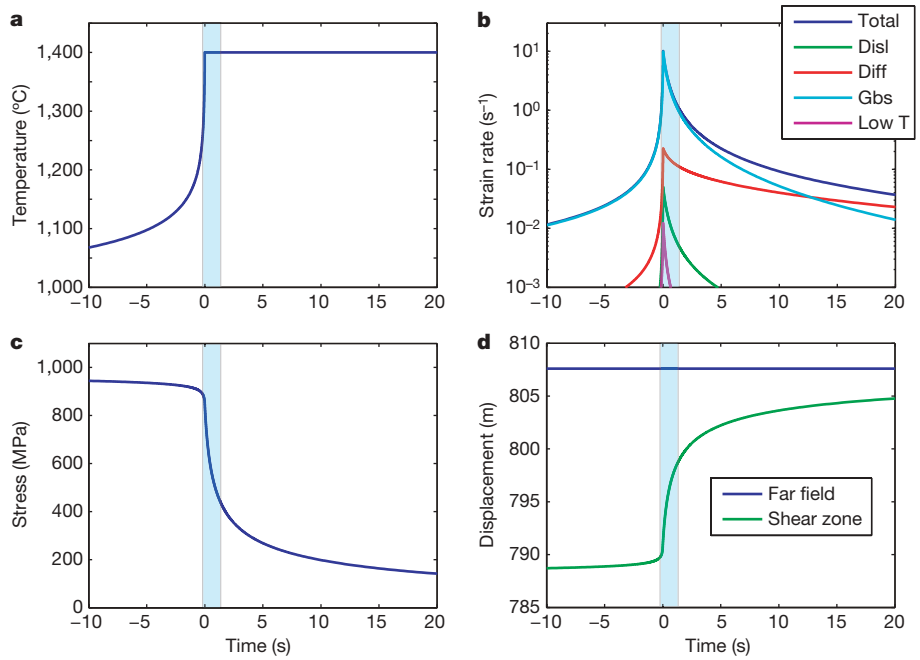


Figure 3 | Same model run as in Figs 2, 4 and Supplementary Fig. S1, but for a 30 s time interval centred on the maximum heating rate in the shear heating event at ~8,040 years. The light blue rectangle highlights the narrow time interval when strain rates are greater than 1 s^{-1}. **a**, Time versus temperature. **b**, Time versus strain rates. 'Total', the sum of strain rates from all four flow laws; 'Disl', dislocation creep (Supplementary equation S1); 'Diff', diffusion creep (Supplementary equation S2); 'Gbs', grain boundary sliding (Supplementary equation S3), 'Low T', low-temperature plasticity (Supplementary equation S4). **c**, Time versus stress. **d**, Time versus displacement.

Equation (3) yields two sets of stable solutions (Fig. 4), one at high stress and low temperatures and another for low stress and unrealistically high temperatures. Steady-state stress and strain can also be calculated for constant temperature. At 1,400 °C, steady-state stress is proportional to strain rate to the sixth power. The arrows in Fig. 4a depict how the system jumps between low and high temperature stable states. The non-dimensional Whitehead number:

$$W = (T_{ss} - T_\infty)/T_\infty = \sigma \dot{\varepsilon} w_{sz} w_{dz}/(C_p \rho \kappa T_\infty) \quad (4)$$

relates T_{ss} to far-field temperature, T_∞. When the Whitehead number is ~1, shear heating has a significant effect on steady-state temperature and strain rate, and the shear zone jumps from the low- to the high-temperature regime (Supplementary Fig. S3).

Steady-state solutions are compared to numerical model results in Fig. 4b and c. For initial model temperature T_0 of 650 °C, evolution of the model shear zone from 650 to 670 °C occurs at low strain rates with slow heating, close to steady state. However, because the velocity along the moving wall of the shear zone is smaller than the far-field velocity, stress rises. Above 670 °C, rapid heating and acceleration of strain occurs, until the maximum temperature of 1,400 °C is attained. At high temperature and high strain rates, model results again approximate steady state. However, in this case, the velocity of the model shear-zone boundary is greater than the far-field velocity, so the stress decreases rapidly. As the stress decreases, the shear zone cools, followed by a jump to a near-steady state at a new, slightly higher T_0.

Our results establish the possibility of repeated earthquake generation via viscous shear heating. Nucleation sites for the instability may originate during earthquakes at oceanic transforms and the outer rise of subduction zones[10,29]. Thus, pre-existing fine-grained

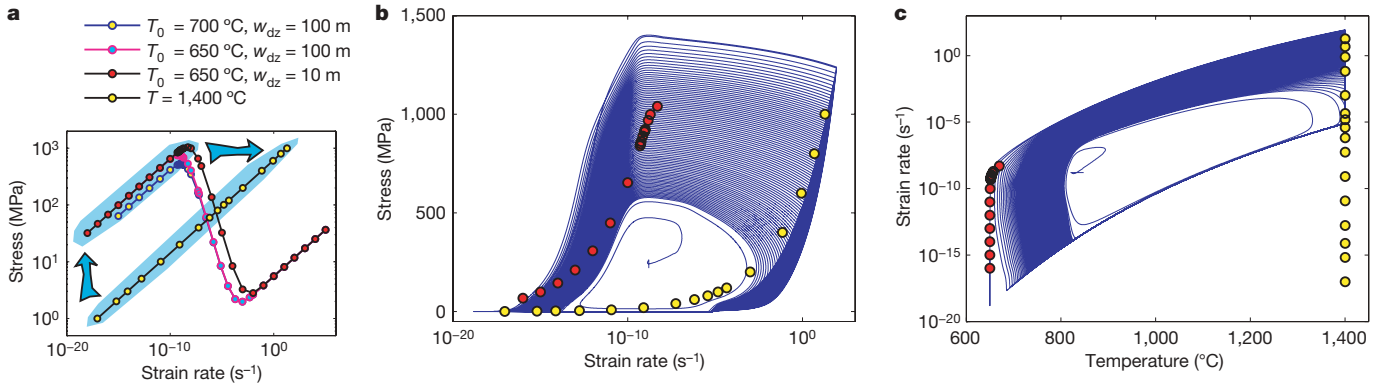


Figure 4 | Results of steady-state analysis, with T_0 = 650 °C and 700 °C, $w_{sz} \approx$ 1 m and $w_{dz} \approx$ 10 and 100 m. **a**, Steady-state strain rate versus stress. The light blue field in the upper left encloses stable solutions for low-temperature steady states. The black line joins steady-state solutions (yellow circles) for a constant temperature of 1,400 °C. Large arrows schematically illustrate the 'limit cycle' followed by a shear zone at a close approach to steady state with an initial temperature of 650 to 700 °C and a maximum temperature of 1,400 °C. Supplementary Fig. S3 illustrates steady-state temperature, strain rate, and stress versus Whitehead number. When the Whitehead number passes through ~1, temperature, strain rate and stress jump from low to high temperature values. **b**, **c**, Plots of strain rate versus stress, and temperature versus strain rate, comparing the numerical results from the model run illustrated in Fig. 2 (blue line) to the predicted steady-state values from 650 to 670 °C (red circles) and at 1,400 °C (yellow circles).

zones have dimensions appropriate for intermediate depth earthquakes as well. In the models, slip velocities greater than $1\,m\,s^{-1}$ occur in the shear zone for several seconds. As shown in the Supplement, stress drops during heating events in our model are limited by the rate of elastic relaxation because the hot, fine-grained shear zone essentially has no strength, and far-field displacement is negligible during these short events.

An intriguing feature of intermediate-depth earthquakes is that they occur mainly between 600 and 800 °C (refs 2, 4, 9, 12). The shear heating mechanism described in this paper operates only within this restricted range of temperature. Above 850 °C, the shear zone deforms at steady-state. Below 600 °C, the stress required to drive the instability in our numerical model is very high (>1.6 GPa). This is far below the brittle fracture strength at pressures of 4 to 6 GPa (ref. 30). However, at such high stresses, grain-size-independent, low-temperature plasticity[25] would distribute deformation over a broad region. We conclude that periodic viscous shear heating instabilities can account for intermediate depth earthquakes within the narrow temperature interval in which they are observed.

Received 8 November 2006; accepted 26 February 2007.

1. Kirby, S. H., Engdahl, E. R. & Denlinger, R. P. in *Subduction Top to Bottom* (eds Bebout, G. E., Scholl, D. W., Kirby, S. H. & Platt, J. P.) 195–214 (Geophysical Monograph 96, American Geophysical Union, Washington DC, 1996).
2. Peacock, S. Are the lower planes of double seismic zones caused by serpentine dehydration in subducting oceanic mantle? *Geology* **29,** 299–302 (2001).
3. Raleigh, C. B. Tectonic implications of serpentinite weakening. *Geophys. J. R. Astron. Soc.* **14,** 113–118 (1967).
4. Hacker, B. R., Peacock, S. M., Abers, G. A. & Holloway, S. D. Subduction factory 2: Are intermediate-depth earthquakes in subducting slabs linked to metamorphic dehydration reactions? *J. Geophys. Res.* **108,** doi:10.1029/2001JB001129 (2003).
5. Jung, H., Green, H. W. & Dobrzhinetskaya, L. F. Intermediate-depth earthquake faulting by dehydration embrittlement with negative volume change. *Nature* **428,** 545–549 (2004).
6. Igarashi, T., Matsuzawa, T., Umino, N. & Hasegawa, A. Spatial distribution of focal mechanisms for interplate and intraplate earthquakes associated with the subducting Pacific plate beneath northeastern Japan arc: A triple-planed deep seismic zone. *J. Geophys. Res.* **106,** 2177–2191 (2001).
7. Kelemen, P. B. & Dick, H. J. B. Focused melt flow and localized deformation in the upper mantle: Juxtaposition of replacive dunite and ductile shear zones in the Josephine peridotite, SW Oregon. *J. Geophys. Res.* **100,** 423–438 (1995).
8. Jaroslow, G. E., Hirth, G. & Dick, H. J. B. Abyssal peridotite mylonites: Implications for grain-size sensitive flow and strain localization in the oceanic lithosphere. *Tectonophysics* **256,** 17–37 (1996).
9. Newman, J., Lamb, W. M., Drury, M. R. & Vissers, R. L. M. Deformation processes in a peridotite shear zone: Reaction softening by an H_2O-deficient, continuous net transfer reaction. *Tectonophysics* **303,** 193–222 (1999).
10. Christensen, D. H. & Ruff, L. Seismic coupling and outer rise earthquakes. *J. Geophys. Res.* **93,** 13421–13444 (1988).
11. Savage, J. C. The mechanics of deep-focus faulting. *Tectonophysics* **8,** 115–127 (1969).
12. Warren, J. & Hirth, G. Grain size sensitive deformation mechanisms in naturally deformed peridotites. *Earth Planet. Sci. Lett.* (submitted).
13. Evans, B., Renner, J. & Hirth, G. A few remarks on the kinetics of static grain growth in rocks. *Int. J. Earth Sci.* **90,** 88–103 (2001).
14. Obata, M. & Karato, S.-I. Ultramafic pseudotachylite from the Balmuccia peridotite, Ivrea-Verbano zone, northern Italy. *Tectonophysics* **242,** 313–328 (1995).
15. Kanamori, H., Anderson, D. L. & Heaton, T. H. Frictional melting during the rupture of the 1994 Bolivian earthquake. *Science* **279,** 839–842 (1998).
16. Whitehead, J. A. & Gans, R. F. A new, theoretically tractable earthquake model. *Geophys. J. R. Astron. Soc.* **39,** 11–28 (1974).
17. Hobbs, B. E., Ord, A. & Teyssier, C. Earthquakes in the ductile regime? *Pure Appl. Geophys.* **124,** 309–336 (1986).
18. Ogawa, M. Shear instability in a viscoelastic material as the cause of deep focus earthquakes. *J. Geophys. Res.* **92,** 13801–13810 (1987).
19. Karato, S.-I. Rheological structure and deformation of subducted slabs in the mantle transition zone; implications for mantle circulation and deep earthquakes. *Phys. Earth Planet. Inter.* **127,** 83–108 (2001).
20. Green, H. W. & Marone, C. in *Plastic Deformation of Minerals and Rocks* (eds Bercovici, D. & Karato, S.) 181–199 (Rev. Mineral. Geochem. 51, Mineralogical Society of America & the Geochemical Society, Washington DC, 2002).
21. Bercovici, D. & Karato, S.-I. in *Plastic Deformation of Minerals and Rocks* (eds Karato, S. & Wenk, H. R.) 387–421 (Rev. Mineral. Geochem. 51, Mineralogical Society of America & the Geochemical Society, Washington DC, 2002).
22. Branlund, J. M., Kameyama, M. C., Yuen, D. A. & Kaneda, Y. Effects of temperature-dependent thermal diffusivity on shear instability in a viscoelastic zone: Implications for faster ductile faulting and earthquakes in the spinel stability field. *Earth Planet. Sci. Lett.* **182,** 171–185 (2000).
23. Kameyama, M. C., Yuen, D. A. & Fujimoto, H. The interaction of viscous heating with grain-size dependent rheology in the formation of localized slip zones. *Geophys. Res. Lett.* **168,** 159–162 (1997).
24. Kameyama, M. C., Yuen, D. A. & Karato, S.-I. Thermal-mechanical effects of low temperature plasticity (the Peierls mechanism) on the deformation of a viscoelastic shear zone. *Earth Planet. Sci. Lett.* **168,** 159–162 (1999).
25. Regenauer-Lieb, K. & Yuen, D. A. Modeling shear zones in geological and planetary sciences: Solid- and fluid-thermal-mechanical approaches. *Earth Sci. Rev.* **63,** 295–349 (2003).
26. Hirth, G. & Kohlstedt, D. in *Inside the Subduction Factory* (ed. Eiler, J.) 83–105 (Geophysical Monograph 138, American Geophysical Union, Washington DC, 2003).
27. Goetze, G. The mechanisms of creep in olivine. *Phil. Trans. R. Soc. Lond. A* **288,** 99–119 (1978).
28. van der Wal, D., Chopra, P., Drury, M. R. & Fitz-Gerald, J. Relationships between dynamically recrystallized grain size and deformation conditions in experimentally deformed olivine rocks. *Geophys. Res. Lett.* **20,** 1479–1482 (1993).
29. Jiao, W., Silver, P. G., Fei, Y. & Prewitt, C. T. Do intermediate- and deep-focus earthquakes occur on preexisting weak zones? An examination of the Tonga subduction zone. *J. Geophys. Res.* **105,** 28125–28138 (2000).
30. Shimada, M., Cho, A. & Yukutake, H. Fracture strength of dry silicate rocks at high confining pressures and activity of acoustic emission. *Tectonophysics* **96,** 159–172 (1983).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Two colleagues were instrumental in helping with this paper: J. Whitehead guided us toward a steady state, and M. Spiegelman helped with a faster thermal diffusion code. In addition we gratefully acknowledge discussions with E. Coon, A. Rubin, P. Molnar, D. McKenzie, M. Billen, J. Gaherty, J. McGuire, L. Montesi, S. Kirby, B. Hacker and J. Warren. This work was supported, in part, by several NSF research grants, the Charles Francis Adams Chair at WHOI (P.B.K.), the Arthur D. Storke Chair at Columbia University (P.B.K.), and a Fellowship from the WHOI Deep Ocean Exploration Institute (G.H.).

Author Contributions While learning from G.H. about the weak fault controversy in Oman, P.B.K. proposed the possibility of a periodic shear heating instability in an upper mantle shear zone of fixed width. P.B.K. constructed the numerical model, and devised the analytical approximation. G.H. provided essential insight on rock mechanics, supplying formulations for olivine flow laws, grain size evolution, stress variation with and without inertial terms, elastic relaxation, and references to prior work on all these topics. G.H. proposed applying the model to intermediate-depth earthquakes. Both authors contributed equally to evaluating and extending model results within natural parameter ranges for temperature, stress, grain size, shear-zone width, and so on, based on our ongoing joint field work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.K. (peterk@ldeo.columbia.edu).

LETTERS

Grassland species loss resulting from reduced niche dimension

W. Stanley Harpole[1] & David Tilman[2]

Intact ecosystems contain large numbers of competing but coexisting species. Although numerous alternative theories have provided potential explanations for this high biodiversity, there have been few field experiments testing between these theories. In particular, theory predicts that higher diversity of coexisting competitors could result from greater niche dimensionality[1], for example larger numbers of limiting resources or factors. Alternatively, diversity could be independent of niche dimensionality because large numbers of species can coexist when limited by just one or two factors if species have appropriate trade-offs[2]. Here we show that plant coexistence and diversity result from the 'niche dimensionality' of a habitat. Plant species numbers decreased with increasing numbers of added limiting soil resources (soil moisture, nitrogen, phosphorus and base cations), which is consistent with theoretical predictions that an increased supply of multiple limiting resources can reduce niche dimension. An observational field study gave similar results. The niche dimension hypothesis also explained diversity changes in the classic Park Grass Experiment at Rothamsted. Our results provide an alternative mechanistic explanation for the effects of nutrient eutrophication on the diversity of terrestrial, freshwater and marine ecosystems.

Multiple resource limitation, heterogeneity, and interspecific differences can explain the stable coexistence of numerous competing species[3]. Niche theory predicts that resource-dependent community assembly imposes a stochastic limit on the similarity of coexisting competing species and thus limits the number of species that can coexist along any particular limiting resource trade-off axis[4]. Resource addition can reduce the number of limiting resources, allowing fewer trade-off opportunities and thus fewer coexisting species. Additionally, an increased supply of limiting resources might change the stoichiometry[5,6] or decrease the heterogeneity of limiting resources, thereby decreasing the size of the region of niche space for coexistence without necessarily reducing the number of limiting resources[3,7]. Because species are evolutionarily differentiated with respect to particular limiting resources, changes in the identity of limiting resources can also reduce the effective niche dimension of existing communities because those species may lack the requisite trait differentiation and trade-offs for coexistence when limited by different resources. Environmental change can therefore reduce the niche dimension of particular ecological communities through decreases in the number or heterogeneity of limiting resources or through changes in the identity and stoichiometry of limiting resources[8].

To experimentally test the response of species number to increased supply of multiple limiting resources, we applied a factorial addition of nitrogen, phosphorus, cations (K^+, Mg^{2+} and Ca^{2+}) and water to a grassland at the University of California's Sedgwick Reserve, Santa Ynez, California, USA (see Methods and Supplementary Information). Plant productivity was limited by all four experimental resources (Fig. 1; significant four-way interaction, Supplementary Table S1), and interactions involving N, P and water were significant for species loss. Resources were strongly co-limiting: we found no significant effect of adding any single resource, but plots that received all four resources had significantly greater productivity and fewer species than did plots that received any of the four possible combinations of three resources (Supplementary Table S2).

To test the 'niche dimension' hypothesis, which predicts that decreased numbers of limiting factors should lead to a decreased number of species, we categorized experimental resource additions as the number of limiting resources added to a plot (that is, 0 = control, 1 = addition of any single resource, 2 = addition of any combination of two resources, and so on). An increased number of resources added, with the objective of overcoming resource limitation, led to a nonlinear decrease in species number because of strong co-limitation (3.1 species were lost with four added resources; Fig. 2a). Changes in the supply of multiple limiting resources may therefore have important consequences for biodiversity in ecosystems that are strongly co-limited: diversity may show little response to increases in the supply of a single resource but show catastrophic loss of species with increases in two or more resources.

The addition of multiple soil resources led to the dominance of a single species in our experiment and to strong decreases in light levels, consistent with a decrease in niche dimensionality towards

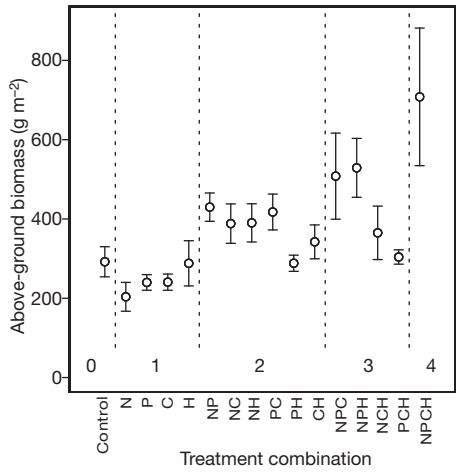


Figure 1 | Multiple resource limitation of plant productivity. The greatest effect of factorial addition of resources occurred with the addition of all four resources: N, P, C and H represent the addition of nitrogen, phosphorus, cations and water. Vertical dotted lines group treatment combinations by the number of added resources, shown at the bottom of each group. Results are shown as means ± s.e.m.

[1]Department of Ecology and Evolutionary Biology, University of California Irvine, 321 Steinhaus Hall, Irvine, California 92697, USA. [2]Department of Ecology, Evolution, and Behavior, University of Minnesota, 100 Ecology Building, 1987 Upper Buford Circle, St Paul, Minnesota 55108, USA.

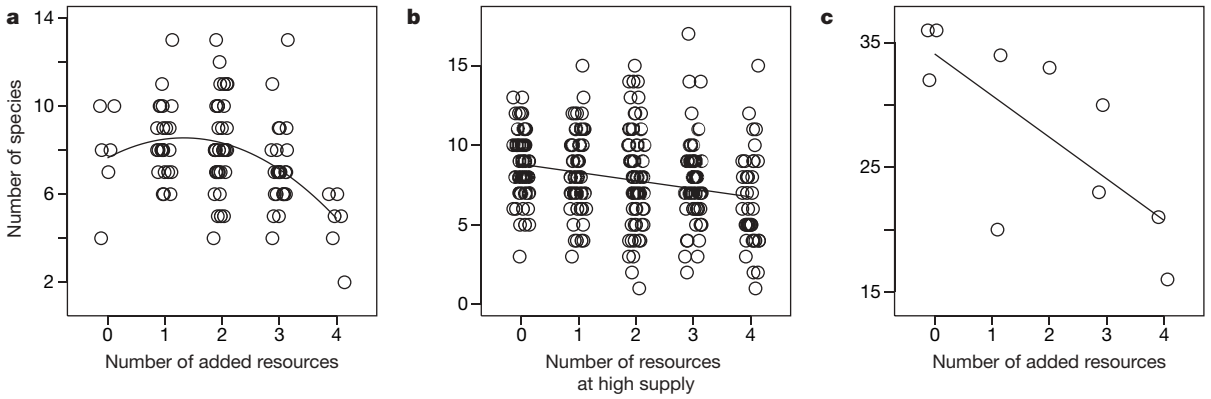


Figure 2 | Species loss with reduction of niche dimension. **a**, Species number plotted against number of added resources ($R^2 = 0.196$, $P < 0.0001$, significant quadratic term, $n = 96$) in the Sedgwick experiment. **b**, Species number plotted against number of resources at high supply ($r = -0.23$, $P = 0.0008$, $n = 215$) in the Sedgwick observational study. **c**, Species number plotted against number of added resources ($R^2 = 0.55$, $P = 0.016$, $n = 10$) in the Park Grass Experiment. Points are jittered for clarity.

single-factor limitation by light and exclusion of all except a single species best adapted to low light but high nutrient concentrations[3]. *Bromus diandrus* increased from 24% average relative biomass in the control plots to 56% in plots receiving all resources. All other species decreased with increasing numbers of added resources (for example *Nassella pulchra* decreased from 57% to 36% relative biomass, and *Bromus hordeaceus* from 13% to 4%).

To compare our experimental results with patterns of species number and nutrient supply in unmanipulated vegetation, we established 215 permanent observational plots in natural grasslands at Sedgwick Reserve. We used buried ion-exchange resins to estimate resource supply for multiple potentially limiting nutrients (N, P, K and Mg; see Methods and Supplementary Table S3). For comparative purposes we categorized the supply of N, P, K and Mg in each observational plot as either high (above the median) or low (below the median) and scored plots by the number (0–4) of these four resources that were at high supply and were therefore less likely to be limiting. Species number decreased with greater numbers of resources at high supply (0.5 species were lost per resource at high supply; Fig. 2b). A loss of less than one species per resource may be because 'high' supply rates reflected a decrease in resource limitation rather than its complete elimination. Although significant, this relationship was weak ($R^2 = 0.053$), indicating the possible importance of other limiting factors besides these particular nutrients as determinants of diversity in these plots.

Our extension of the niche dimension hypothesis is essentially multivariate: species number and productivity both respond directly to the experimental addition of resources (Fig. 3a). The joint response of species number and productivity to the number of resources at high supply was highly significant for the Sedgwick resource addition experiment and observational study (Fig. 3b; multivariate analysis of variance linear contrasts: squared canonical correlations 0.28, 0.086; $F_{2,85} = 16.88$, $F_{2,209} = 9.89$; $P < 0.0001$ and $P < 0.0001$, respectively). Because both species number and evenness determine species diversity, we calculated the 'effective' species number e^H (H is Shannon's diversity index); its value is equivalent to the number of equally abundant species. The diversity (e^H) pattern across the combined natural and augmented productivity gradient at Sedgwick was strikingly monotonic (Fig. 3c), indicating that an increased supply of multiple resources decreased both species number and evenness.

Species loss resulting from fertilization could potentially be a transient response reflecting a local lack of species adapted to a 'vacant niche' represented by the new resource environment[9]. We tested this by using data from the longest-running resource addition experiment of which we are aware[10] (see Methods). The Park Grass Experiment was established at Rothamsted Experimental Station (Harpenden, Hertfordshire, UK) in 1856 in grassland that previously had been mown and grazed for hundreds of years[11]. Since 1856, plots have received annual application of fertilizers[10]. Productivity at this site is limited by multiple resources including N, P and K (ref. 12). Consistent with our experiment at Sedgwick is the observation that increased numbers of added limiting resources led to decreased species number (3.3 species lost per added resource; Fig. 2c). Fertilization has led to species loss that has persisted for 150 years[12], which is consistent with reduced niche dimensionality.

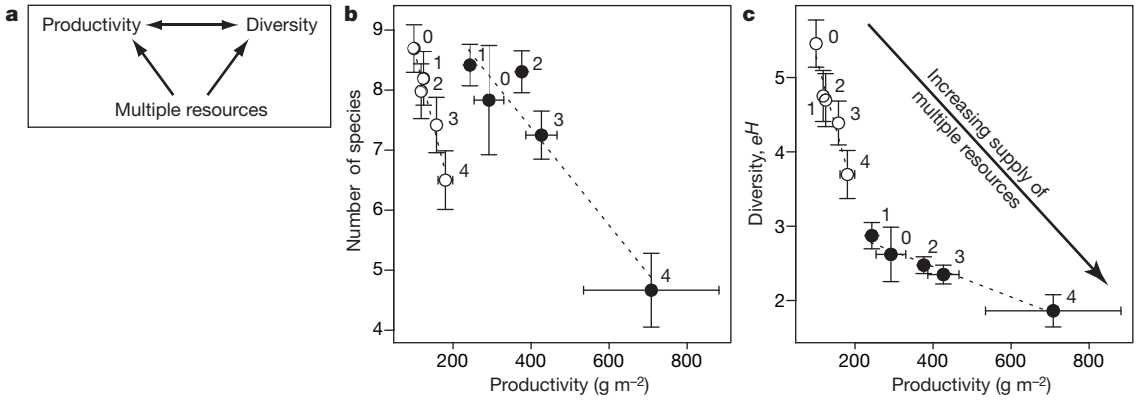


Figure 3 | Effects of multiple resource supply on diversity and productivity across natural and experimental productivity gradients. **a**, Increased supply of multiple resources decreases diversity through reduced niche dimension, and increases productivity; diversity and productivity are correlated. **b**, Joint response of species number and productivity to increased number (labelled numerals) of resources added or at high supply (that is, reduced niche dimension). **c**, Joint response of diversity and productivity to increased number of resources added or at high supply. Open circles in **b** and **c** are observational plots; filled circles are experimental plots. Results are shown as means ± s.e.m.

Although plant species number typically decreases with experimentally increased productivity[10,13–16], it is unclear how productivity could in itself cause species loss. The proposed causal mechanisms of the negative effects of resource addition on species number have been essentially indirect ones mediated through effects of productivity on light and litter[17–19], or through consumers[20,21]. In our experiment, resource addition increased production, and consequently increased litter ($R^2 = 0.13$, $P = 0.0003$) and exponentially decreased light levels ($R^2 = 0.31$, $P < 0.0001$), suggesting possible indirect effects of resource limitation on species number.

To quantify the relative importance of niche dimensionality (the direct effects of the number of limiting resources) as opposed to potential indirect effects of resource addition on species number mediated by productivity, we compared regression models of the number of added resources, light and litter with the use of Akaike's information criteria. We also included productivity as a predictor because it may proximally affect diversity through undetermined mechanisms. Although productivity, but not litter or light, predicted species number, the most parsimonious model (lowest Akaike's information criterion) included only the number of added resources (Supplementary Table S4a). The number of added resources was significant in every model predicting species number when it appeared last, supporting the hypothesis that niche dimensionality contributed to species number even after controlling for potential indirect effects of productivity. Comparing models of e^H, we found that both niche dimension and litter may have been important in determining diversity in this experiment (Supplementary Table S4b).

Our results suggest that a combination of a decreased number of limiting resources and changes in the identity of limiting factors resulting from indirect effects of productivity led to decreased niche dimension and diversity. Decreased resource heterogeneity may also have contributed to decreased niche dimension. Although we did not measure resource heterogeneity in this experiment, in a parallel experiment at Sedgwick the addition of N at 4 g m^{-2} yr^{-1} increased mean inorganic soil N from 2.02 to 7.03 mg kg^{-1} ($P = 0.001$, $n = 16$) and decreased its heterogeneity, expressed as the coefficient of variation, from 0.60 to 0.45. Although the relative importance of these mechanisms is not known, results from the Park Grass Experiment suggest that recovery of diversity after eutrophication may be a slow process. Invasion by preadapted species or evolutionary niche differentiation could eventually restore species diversity, but such consequences are likely to be deemed undesirable or impracticable from a conservation perspective.

Our results suggest that human actions, such as eutrophication, that simplify habitats by decreasing their niche dimensionality can lead to long-term biodiversity loss. Loss of biodiversity could in turn further decrease ecosystem function[22]. Our experimental and observational results from terrestrial grasslands are consistent with those recently found in aquatic systems[23,24]. How species diversity scales with resource diversity in other systems will depend on the degree to which they are co-limited by multiple resources. Experiments are needed that manipulate multiple resources across important environmental gradients such as productivity and latitude in aquatic and terrestrial ecosystems. If niche axes are independent and approximately equal in importance, diversity should decrease linearly with the loss of niche dimension, but the consequences of niche loss may be even greater if species loss responds multiplicatively as suggested[8].

METHODS

We applied N, P, cations (K^+, Mg^{2+} and Ca^{2+}) and water to 96 plots (2 m × 2 m) in a grassland at Sedgwick Reserve, beginning in February 2000. Nutrients were applied twice annually in dry form and water was applied by drip irrigation at roughly one-week intervals throughout the growing season (see Supplementary Information for details and rates). We measured above-ground plant productivity and species composition at peak biomass of the second year of the experiment in May 2001 by clipping, drying and weighing a 0.3-m^2 biomass sample from each plot. The observational plots at Sedgwick consisted of 215 plots (2 m × 2 m) scattered across ungrazed grasslands in the 2,382-ha reserve and encompassed a wide range of topographic variability. We estimated species composition visually and measured total live-plant above-ground biomass by clipping, drying and weighing a 0.2-m^2 sample from each plot. We used inductively coupled plasma chromatography after NaCl extraction of mixed-bed ion-exchange resins that we had buried in the soil in each plot to estimate soil nutrient supply rates over the growing season (see Supplementary Information for details). We used Park Grass species number data from 1946 to 1949 and productivity data from 1949 as reported[10]. Because plant diversity decreased strongly with lower pH caused by the addition of ammonium, we used data from the ten unlimed subplots (plots 2, 3, 4.1, 7, 8, 12, 14, 15, 16 and 17) that did not receive ammonium or manure. Statistical tests are described in more detail in Supplementary Information.

Received 14 October; accepted 12 February 2007.
Published online 25 March 2007.

1. Hutchinson, G. E. Concluding remarks. *Cold Spring Harb. Symp. Quant. Biol.* **22,** 415–427 (1957).
2. Tilman, D. Competition and biodiversity in spatially structured habitats. *Ecology* **75,** 2–16 (1994).
3. Tilman, D. *Resource Competition and Community Structure* (Princeton Univ. Press, Princeton, NJ, 1982).
4. Tilman, D. Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assembly. *Proc. Natl Acad. Sci. USA* **101,** 10854–10861 (2004).
5. Braakhekke, W. G. & Hooftman, D. A. P. The resource balance hypothesis of plant species diversity in grassland. *J. Veg. Sci.* **10,** 187–200 (1999).
6. Herbert, D. A., Rastetter, E. B., Gough, L. & Shaver, G. R. Species diversity across nutrient gradients: An analysis of resource competition in model ecosystems. *Ecosystems* **7,** 296–310 (2004).
7. Chase, J. M. & Leibold, M. A. *Ecological Niches: Linking Classical and Contemporary Approaches* (Univ. Chicago Press, Chicago, IL, 2003).
8. Tilman, D. & Lehman, C. L. Human-caused environmental change: impacts on plant diversity and evolution. *Proc. Natl Acad. Sci. USA* **98,** 5433–5440 (2001).
9. Mittelbach, G. G. *et al.* What is the observed relationship between species richness and productivity? *Ecology* **82,** 2381–2396 (2001).
10. Brenchley, W. E. & Warington, K. *The Park Grass plots at Rothamsted 1856–1949* (Rothamsted Experimental Station, Harpenden, 1958).
11. Silvertown, J. The dynamics of a grassland ecosystem: botanical equilibrium in the Park Grass Experiment. *J. Appl. Ecol.* **17,** 491–504 (1980).
12. Crawley, M. J. *et al.* Determinants of species richness in the Park Grass Experiment. *Am. Nat.* **165,** 179–192 (2005).
13. Lawes, J. B., Gilbert, J. H. & Masters, M. T. Agricultural, botanical and chemical results of experiments on the mixed herbage of permanent meadow, conducted for more than twenty years in succession on the same land. Part II. The botanical results. *Phil. Trans. R. Soc. Lond.* **173,** 1181–1413 (1882).
14. DiTommaso, A. & Aarssen, L. W. Resource manipulations in natural vegetation. *Vegetatio* **84,** 9–29 (1989).
15. Gough, L., Osenberg, C. W., Gross, K. L. & Collins, S. L. Fertilization effects on species density and primary productivity in herbaceous communities. *Oikos* **89,** 428–439 (2000).
16. Baer, S. G., Blair, J. M., Collins, S. L. & Knapp, A. K. Soil resources regulate productivity and diversity in newly established tallgrass prairie. *Ecology* **84,** 724–735 (2003).
17. Goldberg, D. E. & Miller, T. E. Effects of different resource additions on species diversity in an annual plant community. *Ecology* **71,** 213–225 (1990).
18. Foster, B. L. & Gross, K. L. Species richness in a successional grassland: effects of nitrogen enrichment and plant litter. *Ecology* **79,** 2593–2602 (1998).
19. Huston, M. A. & DeAngelis, D. L. Competition and coexistence: the effects of resource transport and supply rates. *Am. Nat.* **144,** 954–977 (1994).
20. Leibold, M. A. Biodiversity and nutrient enrichment in pond plankton communities. *Evol. Ecol. Res.* **1,** 73–95 (1999).
21. Worm, B., Lotze, H. K., Hillebrand, H. & Sommer, U. Consumer versus resource control of species diversity and ecosystem functioning. *Nature* **417,** 848–851 (2002).
22. Loreau, M. *et al.* Biodiversity and ecosystem functioning: Current knowledge and future challenges. *Science* **294,** 804–808 (2001).
23. Grover, J. P. & Chrzanowski, T. H. Limiting resources, disturbance, and diversity in phytoplankton communities. *Ecol. Monogr.* **74,** 533–551 (2004).
24. Interlandi, S. J. & Kilham, S. S. Limiting resources and the regulation of diversity in phytoplankton communities. *Ecology* **82,** 1270–1282 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Chase, J. HilleRisLambers, K. Suding, S. Hobbie, R. Sterner, E. Lonsdorf and G. Oehlert for comments on early drafts of the manuscript, and S. Watts, H. Whitesides and R. Isip for field assistance. We also thank Sedgwick Reserve for research support, and P. Poulton and Rothamsted Research for assistance with data. This work was supported by the Andrew W. Mellon Foundation.

Author Contributions W.S.H. planned and executed the experiment and wrote the manuscript. D.T. contributed to planning and writing.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to W.S.H. (wharpole@uci.edu).

LETTERS

Egalitarian motives in humans

Christopher T. Dawes[1], James H. Fowler[1], Tim Johnson[2,3], Richard McElreath[4] & Oleg Smirnov[5]

Participants in laboratory games are often willing to alter others' incomes at a cost to themselves, and this behaviour has the effect of promoting cooperation[1–3]. What motivates this action is unclear: punishment and reward aimed at promoting cooperation cannot be distinguished from attempts to produce equality[4]. To understand costly taking and costly giving, we create an experimental game that isolates egalitarian motives. The results show that subjects reduce and augment others' incomes, at a personal cost, even when there is no cooperative behaviour to be reinforced. Furthermore, the size and frequency of income alterations are strongly influenced by inequality. Emotions towards top earners become increasingly negative as inequality increases, and those who express these emotions spend more to reduce above-average earners' incomes and to increase below-average earners' incomes. The results suggest that egalitarian motives affect income-altering behaviours, and may therefore be an important factor underlying the evolution of strong reciprocity[5] and, hence, cooperation in humans.

Scarce resources create selective pressure for behaviours that influence how resources are divided in animal societies[6]. When the availability of resources is independent of the choices involved in acquiring them, such behaviours—for example, aggression[6] and begging[7]—can be understood as explicitly motivated by preferences for specific resource divisions. In potentially cooperative encounters, however, resources are produced through organisms' choices to cooperate or defect; thus, behaviours that alter resource allocations produced in cooperative encounters can be viewed as either promoting a cooperation norm or satiating a taste for particular resource divisions[4]. This has sparked considerable debate about how to model social choices[8–10] and has provoked questions[4] concerning the ultimate source of behaviours—such as the reward of contributors[3] and costly punishment of free-riders[1]—that promote cooperation.

For example, in the standard model of multi-person cooperation—the public goods game—cooperation and payoff are correlated. Individuals are endowed with a resource that can be contributed to a common pool; if contributed, the resource increases in value and is divided equally among group members. Social welfare is maximized if all group members contribute, whereas personal wealth is greatest when an individual retains her endowment and others contribute. Past research suggests that individuals are willing to punish those who do not contribute to the common pool[1] and to reward those who do[3]. This behaviour has been interpreted as cooperative norm enforcement[1], but because a player's contribution to the public good is proportional to her payoff from the public good, decreasing the payoff of a defector also has the effect of retrieving economic equality.

Attempts to separate norm enforcement from the pursuit of equality have been incomplete. For instance, one approach[11] alters the efficiency of punishment by making punishment costs equal to the amount punishment reduces incomes. Although this prohibits an individual from reducing inequality between herself and the punished individual, it does not prevent reduction of the standard deviation from the group mean. If a player possessing above average income reduces the income of a wealthier player, then the income difference between high earners and below average earners decreases. Even players with below average income can reduce total inequality if their income is closer to the group mean than the above average earner's income. Thus, even though the inequality between the punisher and the punished player remains the same, punishment can still serve an egalitarian motive in this design[11].

To separate motives, we use a simple experimental design to examine whether individuals reduce or augment others' incomes when there is no cooperative norm to advance (see Methods). We call these behaviours 'taking' and 'giving' instead of 'punishment' and 'reward' to indicate that income alteration cannot change the behaviour of the target. Subjects are divided into groups having four anonymous members each. Each player receives a sum of money randomly generated by a computer. Subjects are shown the payoffs of other group members for that round and are then provided an opportunity to give 'negative' or 'positive' tokens to other players. Each negative token reduces the purchaser's payoff by one monetary unit (MU) and decreases the payoff of a targeted individual by three MUs; positive tokens decrease the purchaser's payoff by one monetary unit (MU) and increase the targeted individual's payoff by three MUs. Groups are randomized after each round to prevent reputation from influencing decisions; interactions between players are strictly anonymous and subjects know this. Also, by allowing participants more than one behavioural alternative, the experiment eliminates possible experimenter demand effects[12]—if subjects were only permitted to punish, they might engage in this behaviour because they believe it is what the experimenters want.

Over the five sessions income alteration was frequent. Among participants, 68% reduced another player's income at least once, 28% did so five times or more, and 6% did so ten times or more. Also, 74% of participants increased another player's income at least once, 33% did so five times or more, and 10% did so ten times or more. Most (71%) negative tokens were given to above-average earners in each group, whereas most (62%) positive tokens were targeted at below-average earners in each group.

The size of income alterations varied with the relative income of the recipient (Fig. 1). Individuals who earned considerably more than other members of their group were heavily penalized. Subjects who earned ten MUs more than the group average received a mean of 8.9 negative tokens compared to 1.6 for those who earned at least ten MUs less than the group. In contrast, individuals who earned considerably less than other group members received sizeable gifts. Subjects who earned ten MUs more than the group average received a mean of 4 positive tokens compared to 11.1 for those who earned at least ten MUs less than the group. Individual spending decisions also suggest that subjects were influenced by concerns for inequality. On average, the bottom earner in each group spent 96% more on

[1]Department of Political Science, University of California, San Diego, California 92093, USA. [2]Center for Adaptive Behaviour and Cognition, Max Planck Institute for Human Development, Lentzeallee 94, Berlin 14195, Germany. [3]Department of Political Science, Stanford University, Palo Alto, California 94305, USA. [4]Department of Anthropology, University of California, Davis, California 95616, USA. [5]Department of Political Science, University of Miami, Coral Gables, Florida 33124, USA.

negative tokens than the top earner and the top earner spent 77% more on positive tokens than the bottom earner (both differences significant, Student's *t*-test, one-tailed, $P < 0.0008$).

Because choices to reduce or to augment others' incomes were costly and yielded no material gain, self-interested subjects had no incentive to engage in it. Those behaviours therefore might decline over time as subjects learn they are not profitable. However, period-specific taking and giving (Fig. 1) shows no consistent pattern over time. Mann–Whitney tests fail to reject the null hypotheses that the number of negative tokens received in period five by above-average earners is the same as that received in periods one to four ($P = 0.38$, two-tailed) or that the number of positive tokens received in period five by below-average earners is the same as that received in periods one to four ($P = 0.86$, two-tailed). Therefore, subjects' income-altering behaviour persists even after acquiring experience playing the game.

To explore how income affects these behaviours, we conduct Tobit regressions of negative and positive tokens received as a function of the positive and negative deviation of one's income from the average income assigned to other group members. We employ robust standard errors clustered on each experimental session. This method allows us to account for the fact that observations are independent only across sessions and that costly taking and giving are censored variables. When examining costly taking, the regression coefficient on 'negative deviation' is –0.45 ($z = -5.11$, $P < 0.001$), and 'positive deviation' is 0.74 ($z = 4.43$, $P < 0.001$); subjects' payoffs are reduced by nearly three-quarters of an MU for each additional MU of income they receive above the average income of other group members. The average income of other group members, when included in the regression, is insignificant ($z = 0.41$, $P = 0.69$). In a model of costly giving, the regression coefficient on 'negative deviation' is 0.83 ($z = 7.56$, $P < 0.001$) and 'positive deviation' is –0.22 ($z = -2.43$, $P = 0.02$); subjects' payoffs are increased by more than eight-tenths of an MU for each additional MU of income below the average income of other group members. The average income of other group members, when included in the regression, is insignificant ($z = -1.41$, $P = 0.18$).

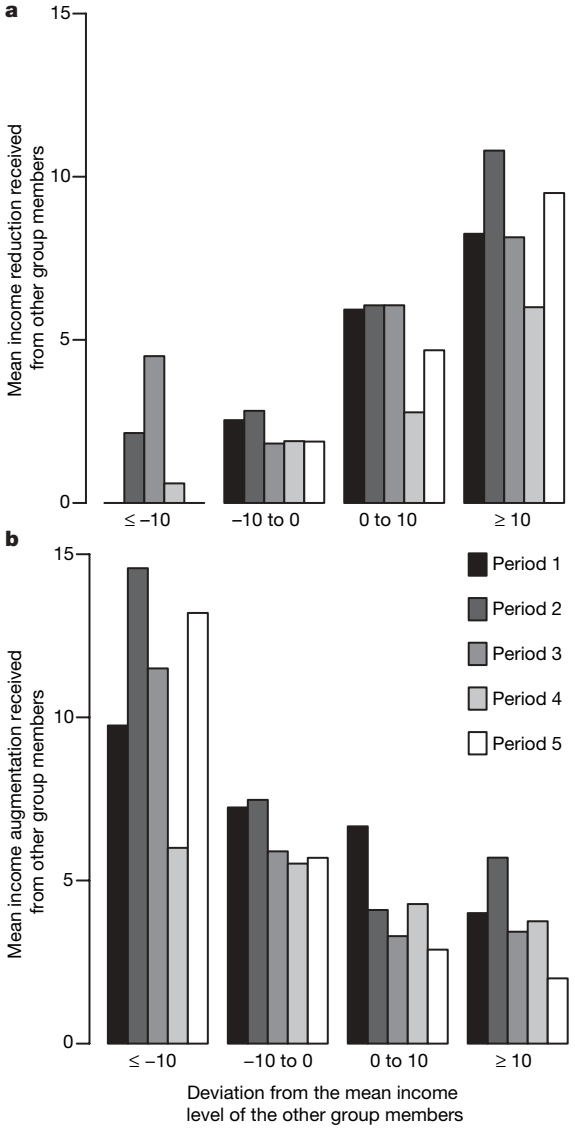


Figure 1 | Mean reduction (a) and augmentation (b) of income by other players in each period as a function of the deviation from the mean income level of the other group members. Income is assigned randomly to each group member by the computer. Reduction and augmentation are costly to the sender—each MU spent on 'costly taking' decreased the recipient's income by three MUs and each MU spent on 'costly giving' increased the recipient's income by three MUs.

We emphasize that income alteration provides no material benefit and, moreover, that a desire for revenge or reimbursement cannot explain choices to reduce or to augment others' incomes. Subjects were told that they never meet the same person twice, so they cannot satisfy, in future rounds, a desire to reciprocate negative or positive tokens assigned to them. To be sure that reciprocation was not a motivation, we conducted additional Tobit regressions. Results show that negative tokens sent were not significantly affected by negative tokens received in the previous round ($z = -0.30$, $P = 0.76$) and positive tokens sent were not significantly affected by positive tokens received in the previous round ($z = -1.17$, $P = 0.24$). Nonetheless, we did observe some behaviours that could not be explained by egalitarian motives. For example, below-average earners sent negative tokens to other below-average earners 12.2% of the time, while below-average earners sent above-average earners positive tokens 16.9% of the time (see Supplementary Information).

In our experiment there is no normative behaviour, so we wondered why people alter incomes. Others[1] show that experimental subjects feel anger towards free-riders in a public goods setting and this anger may motivate punishment. Also, negative emotions inspire the destruction of earned resources when an undeserving party aims to usurp those resources[13] and non-pecuniary expressions of anger satiate the desire to punish individuals who choose not to share a resource equally in experimental games[14]. Income levels are determined by subject behaviour in these experiments, so it is unclear whether resource distributions or anti-social behaviours cause the anger. One possibility is that inequality itself arouses negative emotions. If so, in our experiment we should observe annoyance and anger at high earners; these sentiments should increase as inequality increases and they should be associated with subjects' income-altering behaviours.

To elicit emotional reactions, we presented subjects hypothetical scenarios in which they encountered group members who obtained higher payoffs than they did (see Methods). Subjects were then asked to indicate on a seven-point scale whether they felt annoyed or angry (1, 'not at all'; 7, 'very') by the other individual. In the 'high-inequality' scenario, subjects were told they encountered an individual whose payoff was considerably greater than their own. This scenario generated much annoyance: 75% of the subjects claimed to be at least somewhat annoyed, whereas 41% indicated a high level (4 or more) of annoyance. Many subjects (52%) also indicated that they felt at least some anger towards the top earner. In the 'low-inequality' scenario, differences between subjects' incomes was smaller, and there was significantly less anger (Wilcoxon signed rank test, $P < 0.0001$) and annoyance ($P < 0.0001$). Only 46% indicated they were annoyed and 27% indicated they were angry. Individuals apparently feel negative emotions towards high earners and the intensity of these emotions increases with income inequality.

These emotions seem to influence behaviour. Subjects who said they were at least somewhat annoyed or angry at the top earner in the high-inequality scenario spent 26% more to reduce above-average earners' incomes than subjects who said they were not annoyed or angry. These subjects also spent 70% more to increase below-average earners' incomes. Mann–Whitney tests of both differences indicate that they are significant (one-tailed, $P = 0.05$ and $P = 0.001$, respectively). Emotional reactions towards high earners—even when the source of income is known to be purely random—cause individuals to engage in costly acts that promote equitable resource distributions.

The evidence here indicates that social inequality arouses negative emotions that motivate both the reduction and augmentation of others' incomes. This finding supports research that indicates humans are strongly influenced by egalitarian preferences[7,8]. Furthermore, the results distinguish between models of inequality aversion[8,9]: models that specify which players' incomes will be altered for egalitarian reasons[8] capture subject behaviour better than models that do not[9]. Finally, the results are also consistent with the punishment of non-contributors[1] and the reward of contributors in public good games[3]. Although concerns for equality are clearly not the only motive for human behaviour in these contexts, our results suggest that egalitarian motives may underlie strong reciprocity[11] and, thus, play an important role in the maintenance of cooperation.

METHODS

The design and procedures of the experiment closely approximate a widely cited public good experiment[1]. One hundred and twenty ($n = 120$) students from the University of California at Davis volunteered to participate in the experiment. Recruitment of subjects was conducted in several different departments to maximize the chance that subjects did not know one another; any student who was at least 18 years old was eligible to take part in the study. Twenty subjects attended each of the six experimental sessions and each session involved five periods. Every period, subjects were randomly placed in groups of four subjects. At the beginning of each period subjects received a random payoff and were shown the payoffs for all four members of their group. To maintain comparability with other public good games, random payoffs were drawn from the empirical distribution of payoffs in the first stage of a widely cited public good game with punishment[1]. Subjects were then given an opportunity to either help or harm any member of the group by purchasing up to ten positive and ten negative tokens for each player. At the end of each period, subjects learned the amount of positive and negative tokens they received and their new payoff. The experiment lasted 30 minutes and on average subjects earned approximately ten US dollars per session.

All activity in the experiment was completely anonymous. Group composition changed every period so that no one played with the same person more than once. The subjects were ignorant of other players' experimental history: neither past payoffs nor past decisions were known. Different group composition each period and the absence of any history of play ensured that subjects could neither develop reputations nor target other subjects for revenge.

At the beginning of each session subjects were asked to read experiment instructions on their individual computer screens (see Supplementary Information), and they also had a paper copy available for reference. The instructions explained all features of the experiment, including how payoffs are determined, how group composition is altered every period, and how anonymity of individual decisions and payoffs in the experiment is preserved. In order for the experiment to start, subjects had to answer correctly several test questions designed to ensure full understanding of how choices in the game generate payoffs. At the end of the experimental session, subjects were asked to complete a survey about their demographic characteristics and a questionnaire concerning emotions. The experiment was programmed using GameWeb software written by R.McE.

The emotions questionnaire presented two hypothetical scenarios to subjects: "You receive 23 [19] tokens. The second group member receives 25 [21] and the third 21 [17] tokens. Suppose the fourth member receives 37 [22] tokens. You now accidentally meet this member. Please indicate your feelings towards this person." (Unbracketed numbers were used in the 'high-inequality' scenario and bracketed numbers were used in the 'low-inequality' scenario.) After reading each scenario, subjects were asked to indicate on a seven-point scale whether they felt annoyed or angry (1, 'not at all'; 7, 'very').

Received 8 November 2006; accepted 5 February 2007.

1. Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415,** 137–140 (2002).
2. Andreoni, J., Harbaugh, W. & Vesterlund, L. The carrot or the stick: rewards, punishments, and cooperation. *Am. Econ. Rev.* **93,** 893–902 (2003).
3. Sefton, M., Shupp, R. & Walker, J. The effect of rewards and sanctions in the provision of public goods. *Center for Applied Economics and Policy Research Working Paper* **2006–005,** 1–43 (2006).
4. Fowler, J. H., Johnson, T. & Smirnov, O. Egalitarian motive and altruistic punishment. *Nature* **433,** E1, doi:10.1038/nature03256 (2005).
5. Fehr, E., Fischbacher, U. & Gächter, S. Strong reciprocity, human cooperation, and the enforcement of social norms. *Hum. Nat.* **13,** 1–25 (2002).
6. Maynard Smith, J. & Price, G. The logic of animal conflict. *Nature* **146,** 15–18 (1973).
7. Bergstrom, C. T. & Lachmann, M. Signaling among relatives. III. Talk is cheap. *Proc. Natl Acad. Sci. USA* **95,** 5100–5105 (1998).
8. Fehr, E. & Schmidt, K. M. A theory of fairness, competition, and cooperation. *Q. J. Econ.* **114,** 817–868 (1999).
9. Bolton, G. & Ockenfels, A. ERC: A theory of equity, reciprocity, and competition. *Am. Econ. Rev.* **90,** 166–193 (2000).
10. Falk, A. & Fischbacher, U. A theory of reciprocity. *Games Econ. Behav.* **54,** 293–315 (2006).
11. Falk, A., Fehr, E. & Fischbacher, U. Driving forces behind informal sanctions. *Econometrica* **73,** 2017–2030 (2005).
12. Orne, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *Am. Psych.* **17,** 776–783 (1962).
13. Bosman, R. & van Winden, F. Emotional hazard in a power-to-take experiment. *Econ. J.* **112,** 147–169 (2002).
14. Xiao, E. & Houser, D. Emotion expression in human punishment behaviour. *Proc. Natl Acad. Sci. USA* **102,** 7398–7401 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Center for Adaptive Behaviour and Cognition at the Max Planck Institute for Human Development and the UC Davis Institute of Government Affairs for generous research support.

Author Contributions The authors are listed alphabetically because each author contributed equally to the design, implementation, analysis and communication of this research.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.H.F. (jhfowler@ucsd.edu).

LETTERS

A mirror-symmetric cell division that orchestrates neuroepithelial morphogenesis

Marcel Tawk[1], Claudio Araya[1], Dave A. Lyons[1]†, Alexander M. Reugels[2], Gemma C. Girdler[1], Philippa R. Bayley[1]†, David R. Hyde[3], Masazumi Tada[1] & Jonathan D. W. Clarke[1]

The development of cell polarity is an essential prerequisite for tissue morphogenesis during embryogenesis, particularly in the development of epithelia[1,2]. In addition, oriented cell division can have a powerful influence on tissue morphogenesis[3]. Here we identify a novel mode of polarized cell division that generates pairs of neural progenitors with mirror-symmetric polarity in the developing zebrafish neural tube and has dramatic consequences for the organization of embryonic tissue. We show that during neural rod formation the polarity protein Pard3 is localized to the cleavage furrow of dividing progenitors, and then mirror-symmetrically inherited by the two daughter cells. This allows the daughter cells to integrate into opposite sides of the developing neural tube. Furthermore, these mirror-symmetric divisions have powerful morphogenetic influence: when forced to occur in ectopic locations during neurulation, they orchestrate the development of mirror-image pattern formation and the consequent generation of ectopic neural tubes.

The movements of neurulation in zebrafish embryos involve convergence of left and right sides of the neural plate towards the dorsal midline, followed by invagination of the neural plate to form a neural keel that then condenses to form a solid neural rod[4–7]. At the midline of the neural rod, an epithelial seam forms that divides left from right and initiates formation of the ventricular system of the brain and spinal cord[8]. The neural keel and rod stages are uniquely characterized by neural progenitor divisions that deposit one daughter cell on either side of the midline, so that each side of the neural tube receives contributions from both left and right sides of the neural plate[4–7,9] (Supplementary Fig. 1). Once deposited on either side of the neural rod, both daughter cells elongate across the apico-basal extent of the left and right neuroepithelium (Fig. 1a). More than 90% of neural plate cells undergo this midline-crossing division (C-division)[10]. When cell division is blocked during this period, very few cells are able to cross the midline[7] (Fig. 1b–e), thus demonstrating that division itself is required for crossing. The daughter cells that cross the midline integrate into a neuroepithelium that develops mirror-image apico-basal polarity with respect to their original side. To integrate into the contra-lateral neuroepithelium, a mechanism must exist to generate mirror-image polarity in the crossing daughter cells.

C-divisions are most prevalent at neural keel and rod stages (14 to 18 hours post fertilization (h.p.f.))[6,9], which coincides with the first midline expression of markers of apical epithelial character[9]. Before these stages, apical markers are either not expressed (aPKC) or are expressed diffusely in neural plate cell membranes (β-catenin and ZO-1)[9]. Because Par3 has a fundamental role in the establishment of cell polarity in a variety of invertebrate and vertebrate species[11], we chose to follow the dynamics of tissue and cell polarization using time-lapse confocal microscopy of a fusion protein of Pard3 with green fluorescent protein, Pard3–GFP (Pard3, initially described as ASIP/PAR-3, is a zebrafish orthologue of the *Caenorhabditis elegans* Par3 protein[9,12,13]). A few evenly distributed puncta of Pard3–GFP are observed in the neural plate[9], but the first coordinated localization appears towards the end of the neural keel phase when Pard3–GFP is expressed in bright puncta within a zone approximately 40 μm wide at the midline (Fig. 2b). These puncta are mobile and gradually coalesce towards the midline over the next few hours. By 18 h.p.f., a distinct single midline expression domain that extends throughout the dorso-ventral extent of the neural rod is evident (Fig. 2c; and Supplementary Fig. 2a, b).

When expression in individual cells is examined, the development of bright Pard3–GFP puncta is particularly striking in cells undergoing the C-division. When cell bodies round up at the start of prophase, Pard3–GFP has no consistent subcellular localization, however, by telophase, Pard3–GFP is specifically enriched at the cleavage furrow (Fig. 2d; and Supplementary Fig. 2). In most cells, bright expression remains symmetrically localized across the cleavage

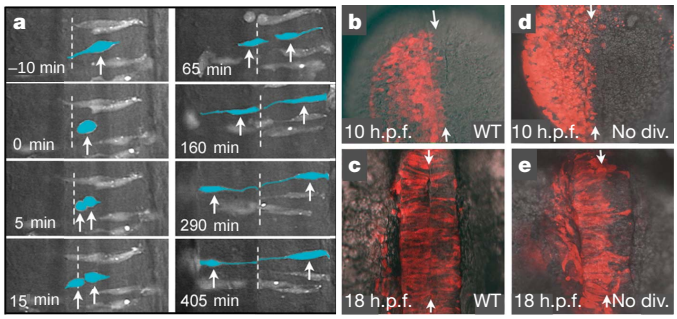


Figure 1 | Cell division separates daughter cells across the midline of the neural keel and rod. **a**, Time-lapse sequence of a C-division (crossing-division) viewed from the dorsal surface of a neural rod. The cell divides on the right side of the neural rod and is followed by the medial daughter crossing the midline (dotted line) and integrating into the contra-lateral side. Mother and daughter cells are highlighted by arrows and blue overlay. Time points are in minutes relative to the beginning of mitosis. **b**, Dorsal view of a normal neural plate with cells labelled by red fluorescence on only the left side. Arrows point to the midline. **c**, Dorsal view of the same embryo as **b** but 8 h later, showing bilateral distribution of red cells. **d**, Dorsal view of a neural plate in an embryo treated with inhibitors of cell division (No div.) during the period of C-divisions. Red fluorescent cells are predominantly on the left side of the neural plate. **e**, Dorsal view of the neural keel 8 h later than **d**, indicating that very few cells crossed the midline (arrows) in the absence of cell division. WT, wild type.

[1]Anatomy and Developmental Biology, UCL, Gower Street, London WC1E 6BT, UK. [2]Institut für Entwicklungsbiologie, Universität zu Köln, 50923 Köln, Germany. [3]Department of Biological Sciences and Center for Zebrafish Research, University of Notre Dame, Notre Dame, Indiana 46556, USA. †Present addresses: Department of Developmental Biology, Stanford University School of Medicine, Beckman Center B300, 279 Campus Drive, Stanford, California 94305, USA (D.A.L.); Department of Academic and Student Affairs, Oregon Health and Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239, USA (P.R.B).

furrow throughout cytokinesis (40 out of 47 monitored divisions; Fig. 2e). In a few cells, Pard3–GFP was only enriched at the cleavage plane towards the end of cytokinesis (7/47). Intense expression of Pard3–GFP remains at the medial ends of the two daughter cells following all C-divisions (Fig. 2e). During this period, the most medial daughters move across the midline and elongate to stretch across the full prospective apico-basal extent of the contra-lateral epithelium, while their sister cells elongate across the full extent of the ipsi-lateral epithelium. Thus, mirror-image polarity of daughter cells is established during cytokinesis by a novel mechanism that localizes apical information to the cleavage furrow.

To test whether Pard3 is important for establishing mirror-image polarity during C-divisions and for initiating midline crossing of the medial daughters, we first reduced Pard3 levels by morpholino-mediated antisense knockdown (Supplementary Fig. 3). Reduction of Pard3 had no effect on the frequency, location or completeness of cell divisions (Supplementary Fig. 4), but did dramatically decrease midline crossing in 7/10 embryos (Fig. 2f). We next tested whether a mutant version of Pard3 that does not preferentially localize to the cleavage furrow during the C-division (97/102 cells monitored; Fig. 2g) would inhibit midline crossing. We mosaically expressed the mutant Pard3-Δ6–GFP, which lacks amino acids 688–1127 including the aPKC binding domain[13]. In neuroepithelial cells, this protein is expressed widely in the cell membrane rather than being restricted to apical poles and also associates with intracellular filaments (probably microtubules)[13]. Time-lapse analyses revealed that for Pard3–GFP-expressing cells 80% (32/40) of cell divisions led to midline crossing less than 10 min after prophase, and 100% cross by 35 min. In contrast, only 49% (39/80) of daughter cells expressing Pard3-Δ6–GFP cross the midline within 10 min, and in the remaining 51% (41/80) of divisions, both daughters stay on the same side until lost from focus (Fig. 2g and Supplementary Table 1). Together, these results demonstrate that Pard3 function is required for midline crossing and indicate that this role for Pard3 is probably mediated by specific localization to the cleavage plane and mirror-symmetric inheritance at the prospective apical poles of the daughter cells.

To investigate the regulation of the C-division, we looked for mirror-symmetric divisions in embryos in which convergence of the neural plate and formation of the neural keel was delayed by reducing the function of the non-canonical Wnt signalling component Vangl2 (refs 14–16; Fig. 3a, b). We reasoned that C-divisions would occur 'on time' but in ectopic lateral locations, if regulated by a mechanism intrinsic to the neuroepithelial precursors, but would occur 'late' in the delayed neural keel, if regulated by the process of neurulation or the keel environment. We find the former is true. Time-lapse analyses in the region of the caudal hindbrain and anterior spinal cord reveal that ectopic C-divisions occur in neural plate cells before keel formation in *trilobite/vangl2* mutants (Fig. 3c, d; and Supplementary Movies 1 and 2). These divisions occur across the deep-to-superficial axis of the convergence-delayed neural plates and generate pairs of daughter cells that remain in register as they

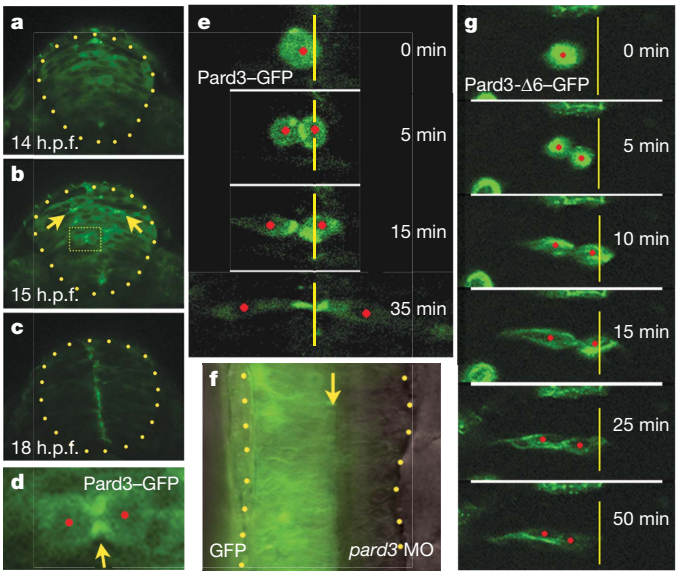


Figure 2 | Subcellular distribution of Pard3–GFP reveals that C-division is a mirror-symmetric division. **a**–**c**, Three frames from a time-lapse sequence of Pard3–GFP expression in neural keel and rod seen in transverse section. In early keel, Pard3–GFP is diffusely expressed throughout the cytoplasm of cells and shows no polarization. At approximately 15 h.p.f., bright puncta appear within a territory 20 μm either side of the midline and by 18 h.p.f. these puncta accumulate at the midline of the rod and the diffuse cytoplasmic distribution has disappeared. **d**, High magnification of a telophase cell outlined in **b** shows distinct Pard3–GFP accumulation around the cleavage furrow (arrow). **e**, A time-lapse sequence showing Pard3–GFP distribution throughout a C-division. Pard3–GFP is distributed across the cleavage furrow and inherited in medial poles of the two daughter cells. The medial daughter rapidly crosses the midline (yellow line). Time points are in minutes, starting at the beginning of mitosis. **f**, Dorsal view of a neural tube in which the level of Pard3 was reduced by injection of 0.35 pmoles of morpholino (MO). The unilateral distribution of GFP-labelled cells indicates that few cells crossed the midline (arrow), compared to the normal embryo in Fig. 1c. **g**, Time-lapse sequence of a mutant Pard3-Δ6–GFP-expressing cell dividing close to the neural keel midline (yellow line). Pard3-Δ6–GFP expression is not localized to the cleavage furrow or to prospective apical poles of the daughter cells. Both daughter cells remain on the same side of the keel for at least 50 min before drifting out of focus. Time points are in minutes, starting at the beginning of mitosis.

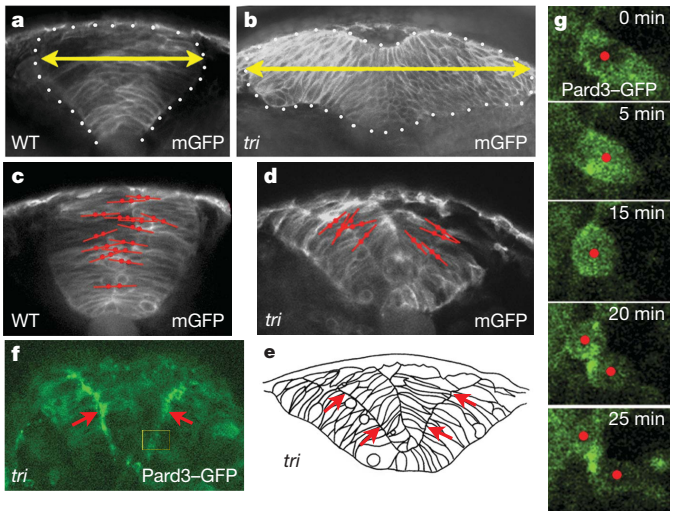


Figure 3 | Delayed convergence of the neural plate leads to ectopic Pard3-mediated C-divisions. **a**, Arrowheads indicate the width of wild-type neural keel in live confocal transverse sections at 14 h.p.f. Cells are labelled with membrane-bound GFP (mGFP). **b**, The *trilobite/vangl2* mutant neural plate, also at 14 h.p.f., has a much wider profile owing to delayed convergence. **c**, The orientations of cell divisions are superimposed on wild-type neural rod. Red lines and dots indicate direction of separation of daughter cells, leading to midline crossing. **d**, Positions and orientations of cell divisions towards the end of the invagination of *trilobite/vangl2* (*tri*) neural plate. Daughters are separated across what was the superficial-deep axis of the plate. **e**, Line drawing of cell-outlines traced from the *trilobite/vangl2* specimen illustrated in **d** to show ectopic interface (arrows) between developing bilayers in mutant neural primordium. **f**, A section of *trilobite/vangl2* neural primordium at 18 h.p.f. reveals two ectopic, bilateral lines of Pard3–GFP expression (arrows) coincident with the ectopic bilateral interface between cells illustrated in **e**. **g**, Time-lapse sequence showing Pard3–GFP expression concentrated at the cleavage furrow of a cell dividing across the neural plate in a *trilobite/vangl2* embryo. Location of the division is shown by the boxed area in **f**. Time points are in minutes, starting at the beginning of mitosis. Red dots indicate the centre of mother and daughter cells.

elongate across the neural plate (Supplementary Fig. 5). Because the daughter cells remain in register, the neural plate becomes a bilayer of pseudostratified epithelia (Fig. 3e). In contrast to wild-type embryos, in which a single domain of Pard3–GFP expression develops only in the midline of the neural keel (Fig. 2c), Pard3–GFP in *trilobite/vangl2* mutants is expressed in two domains on either side of the embryo's midline, sandwiched midway between the superficial and deep surfaces of the neural plate (Fig. 3f). Examination of individual cells reveals that Pard3–GFP is mirror-symmetrically expressed across the cleavage plane of cells that divide across the mutant bilayered neural plate (30/37 cells; Fig. 3g; and Supplementary Movies 3 and 4). These observations all suggest that ectopic neural 'midlines' are being generated by ectopic C-divisions on either side of the actual embryonic midline, and that Pard3-mediated C-divisions are independent of non-canonical Wnt function (see also Supplementary Fig. 6).

Despite the abnormal development of the neural plate, neural tissue continues to converge and invaginate in *trilobite/vangl2* mutant embryos. Thus, by 24 h.p.f., the ectopically bilayered neural plates are re-oriented and appear to fuse at the midline, to generate a four-layered neural primordium. Molecular and structural analyses show this four-layered structure comprises two neural tubes side-by-side. The apical marker aPKC, which normally delineates the single neural midline (Fig. 4a), is expressed along the two ectopic neural midlines in *trilobite/vangl2* mutants (Fig. 4d). Remarkably, GFAP

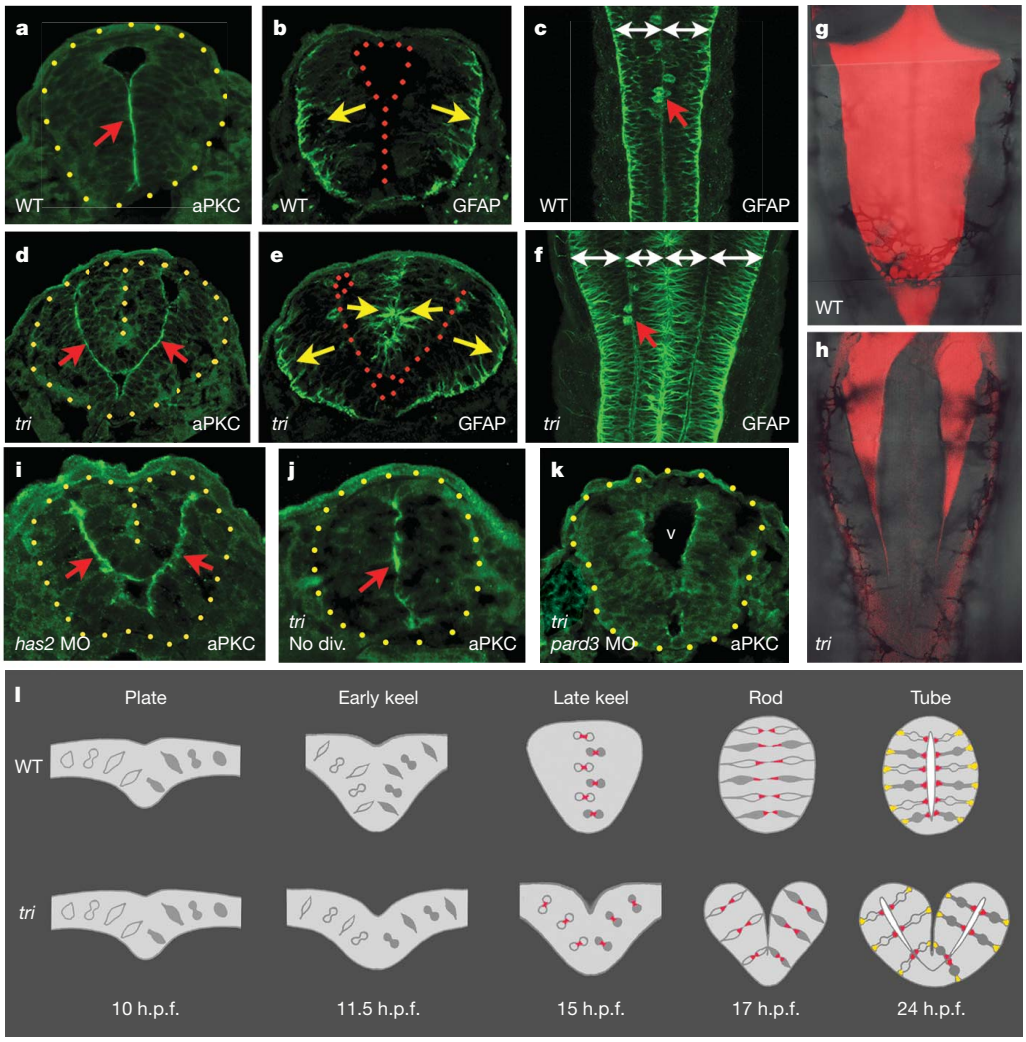


Figure 4 | Delayed neural plate convergence generates duplicate neural tubes complete with mirror-image apico-basal polarity and ventricles. **a**, A transverse section of wild-type neural tube at the level of the caudal hindbrain at 24 h.p.f. reveals aPKC expression at the apical surface (red arrow) of the epithelium. Yellow dots indicate the basal surface of the tube. **b**, A transverse section of neural tube reveals GFAP expression at the basal edge (yellow arrows) of epithelia. Red dots indicate the apical surfaces of the tube. **c**, A horizontal confocal section of a wild-type neural tube whole-mount stained for GFAP expression. As well as marking basal ends of neuroepithelial cells this antigen is highly expressed in mitotic cells (red arrow) at the apical surface. **d**, **e**, A transverse section of *trilobite/vangl2* neural primordium at 24 h.p.f. reveals bilateral aPKC expression (two red arrows) and mirror-image expression of GFAP (four yellow arrows) on either side of apical surfaces. **f**, A horizontal confocal section of *trilobite/vangl2* neural tube stained as a whole-mount for GFAP expression reveals duplicate mirror-image epithelial organization (compare to **c**). **g**, Fluorescent dextran injected into a ventricle of a wild-type embryo reveals a single, large ventricle in the hindbrain. **h**, Fluorescent dextran injected into a ventricle of a *trilobite/vangl2* embryo reveals twinned ventricles in the caudal hindbrain. **i**, A transverse section of neural primordium in a *has2* morpholino embryo at 24 h.p.f. reveals bilateral aPKC expression that phenocopies the *trilobite/vangl2* phenotype. **j**, A transverse section of neural tube from a *trilobite/vangl2* embryo treated with inhibitors of cell division. Epithelial duplications are significantly reduced in 76% (32/42) of mutant embryos as revealed by single midline expression of the apical marker aPKC (compare to **d**). **k**, A transverse section of neural tube from a *trilobite/vangl2* embryo injected with 0.35 pmoles of *pard3* morpholino oligonucleotides. Epithelial duplications are significantly reduced in 46% of mutant embryos as revealed by the single midline ventricle (v) and expression of the apical marker aPKC at the ventricular surface. **l**, A schematic of transverse sections to show mirror-symmetric C-divisions during the morphogenesis of normal and mirror-duplicate neural tubes. Pard3 localization is shown in red and the basal marker GFAP is shown in yellow.

which is a marker of basal (outer) surface of the neuroepithelium (Fig. 4b, c) is expressed with mirror-image symmetry on either side of each of the two apical aPKC domains (Fig. 4e, f). By 30 h.p.f., ectopic ventricles with appropriate dorso-ventral morphology are generated on either side of the midline (Fig. 4g, h). Identical phenotypes are obtained by abrogation of other components of the non-canonical Wnt pathway such as Pk1 (ref. 17) and Dvl (refs 18, 19) (Supplementary Fig. 7).

To confirm that generation of mirror-image duplicates is a consequence of delayed convergence rather than other potential defects downstream of non-canonical Wnt signalling (see ref. 7 and Supplementary Fig. 6), we analysed embryos with convergence defects caused either by reduced function of the Hyaluron synthesizing enzyme, Has2, which is in the mesoderm[20], or by surgical separation of the left from right side of the neural plate in wild-type embryos (Supplementary Fig. 8). We find both interventions lead to mirror-image duplication of the neural tubes, identical to embryos with non-canonical Wnt signalling deficits (Fig. 4i; and Supplementary Fig. 8 and Supplementary Movie 5). We conclude that when neural keel formation is delayed or prevented, ectopic mirror-symmetric divisions occur across the neural plate, and we propose that these specialized divisions initiate mirror-image apico-basal organization across the neuroepithelium.

If the ectopic mirror-image apico-basal organization is really a downstream consequence of the ectopic mirror-symmetric cell divisions then blocking division should eliminate this phenotype. In fact a previous study has demonstrated that blocking division during keel formation rescues normal development of the neural tube in *trilobite/vangl2* embryos[7], and we confirm this here (Fig. 4j). In addition, because Pard3 function is crucial for the mirror-symmetric division, we tested whether it is also required for mirror-image duplications by reducing Pard3 levels in *trilobite/vangl2* mutants. We find that a partial reduction of Pard3 levels by morpholino knockdown reduces mirror-image duplication in 46% of the mutants (7/39 were fully rescued, whereas 11/39 showed reduced duplication; Fig. 4k; and Supplementary Fig. 9). More severe reduction of Pard3 function disrupts morphogenesis such that it is impossible to score for duplications. These results demonstrate that ectopic mirror-symmetric divisions and Pard3 function are required to generate duplicated mirror-image neural tubes in embryos with delayed convergence.

A prerequisite for lumen formation in the solid neural rod is the development of apical specializations at the tissue midline and mirror-image cell polarization on either side of this midline. Our results (summarized in Fig. 4l) demonstrate that this tissue organization is orchestrated by mirror-symmetric cell division, which localizes the apical polarity protein Pard3 to the mitotic cleavage furrow. It will be interesting in the future to see if mirror-symmetric divisions are a more general mechanism for initiating mirror-image pattern or lumen formation in other embryonic organ primordia.

METHODS

Embryo care. Embryos were staged and cared for according to standard protocols. The *trilobite/vangl2* mutant allele used was tri^{m209} (ref. 14).

Time-lapse imaging. Confocal time-lapse imaging was carried out by embedding embryos in low melting point agarose (Sigma) and viewing the neural keel in the transverse or dorsal planes using a ×40 or a ×63 long-working-distance water immersion objective. Embryos were maintained at 28.5 °C in an environmental chamber and *z*-stacks collected at 5-min intervals, usually starting at 10 or 11 h.p.f. and continuing through to 18 h.p.f.

Immunocytochemistry. For immunostaining, embryos were fixed in 4% paraformaldehyde and sectioned at 10 μm on a cryostat, or stained as wholemounts. Antibodies and dilutions used were aPKCζ (Santa Cruz C20, 1:500), GFAP (Dako Z0344, 1:200) and phospho-histone H3 (Sigma H0412, 1:200).

Cell division inhibitors. To block cell division at neural keel and rod stages, embryos were cultured in embryo medium containing 100 μM aphidicolin (Sigma) and 20 mM hydroxyurea (Sigma) dissolved in 4% dimethylsulphoxide[7,8,21], from approximately 90% epiboly onwards.

Dextran injection into ventricles. To reveal ventricular organization in live embryos, a small volume of 4% rhodamine dextran was pressure-injected into the midbrain ventricle of wild-type and experimental embryos at 30 h.p.f.

Additional details of methods, including information about the morpholinos and RNA constructs used can be found in Supplementary Information.

Received 31 October 2006; accepted 26 February 2007.
Published online 28 March 2007.

1. Keller, R. Shaping the vertebrate body plan by polarized embryonic cell movements. *Science* **298,** 1950–1954 (2002).
2. Wodarz, A. Establishing cell polarity in development. *Nature Cell Biol.* **4,** 39–44 (2002).
3. Gong, Y., Mo, C. & Fraser, S. E. Planar cell polarity signalling controls cell division orientation during zebrafish gastrulation. *Nature* **430,** 689–693 (2004).
4. Papan, C. & Campos-Ortega, J. A. A clonal analysis of spinal cord development in the zebrafish. *Dev. Genes Evol.* **207,** 71–81 (1997).
5. Papan, C. & Campos-Ortega, J. A. Region-specific cell clones in the developing spinal cord of the zebrafish. *Dev. Genes Evol.* **209,** 135–144 (1999).
6. Kimmel, C. B., Warga, R. M. & Kane, D. A. Cell cycles and clonal strings during formation of the zebrafish central nervous system. *Development* **120,** 265–276 (1994).
7. Ciruna, B., Jenny, A., Lee, D., Mlodzik, M. & Schier, A. F. Planar cell polarity signalling couples cell division and morphogenesis during neurulation. *Nature* **439,** 220–224 (2006).
8. Lowery, L. A. & Sive, H. Initial formation of zebrafish brain ventricles occurs independently of circulation and requires the *nagie oko* and *snakehead/atp1a1a.1* gene products. *Development* **132,** 2057–2067 (2005).
9. Geldmacher-Voss, B., Reugels, A. M., Pauls, S. & Campos-Ortega, J. A. A 90° rotation of the mitotic spindle changes the orientation of mitoses of zebrafish neuroepithelial cells. *Development* **130,** 3767–3780 (2003).
10. Lyons, D. A., Guy, A. T. & Clarke, J. D. Monitoring neural progenitor fate through multiple rounds of division in an intact vertebrate brain. *Development* **130,** 3427–3436 (2003).
11. Macara, I. G. Parsing the polarity code. *Nature Rev. Mol. Cell Biol.* **5,** 220–231 (2004).
12. Wei, X. *et al.* The zebrafish Pard3 ortholog is required for separation of the eye fields and retinal lamination. *Dev. Biol.* **269,** 286–301 (2004).
13. Von Trotha, J. W., Campos-Ortega, J. A. & Reugels, A. M. Apical localization of ASIP/PAR-3:EGFP in zebrafish neuroepithelial cells involves the oligomerization domain CR1, the PDZ domains, and the C-terminal portion of the protein. *Dev. Dyn.* **235,** 967–977 (2006).
14. Jessen, J. R. *et al.* Zebrafish *trilobite* identifies new roles for Strabismus in gastrulation and neuronal movements. *Nature Cell Biol.* **4,** 610–615 (2002).
15. Park, M. & Moon, R. T. The planar cell-polarity gene *stbm* regulates cell behaviour and cell fate in vertebrate embryos. *Nature Cell Biol.* **4,** 20–25 (2002).
16. Goto, T. & Keller, R. The planar cell polarity gene *strabismus* regulates convergence and extension and neural fold closure in *Xenopus. Dev. Biol.* **247,** 165–181 (2002).
17. Carreira-Barbosa, F. *et al.* Prickle 1 regulates cell movements during gastrulation and neuronal migration in zebrafish. *Development* **130,** 4037–4046 (2003).
18. Wallingford, J. B. & Harland, R. M. Neural tube closure requires Dishevelled-dependent convergent extension of the midline. *Development* **129,** 5815–5825 (2002).
19. Tada, M. & Smith, J. C. *Xwnt11* is a target of *Xenopus* Brachyury: regulation of gastrulation movements via Dishevelled, but not through the canonical Wnt pathway. *Development* **127,** 2227–2238 (2000).
20. Bakkers, J. *et al.* Has2 is required upstream of Rac1 to govern dorsal migration of lateral cells during zebrafish gastrulation. *Development* **131,** 525–537 (2004).
21. Lyons, D. A. *et al. erbb3* and *erbb2* are essential for Schwann cell migration and myelination in zebrafish. *Curr. Biol.* **15,** 513–524 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank P. Alexandre, D. Barker, J. Brockes, M. Costa, M. Kai, R. Sousa-Nunes, V. Prince and S. Wilson for comments and discussion on the manuscript; M. Costa for Supplementary Movie 1; S. Goulas for Fig. 4c, f; and M. Hammerschmidt for the *has2* morpholino. This work was funded by the MRC, the BBSRC and the Wellcome Trust.

Author Contributions M. Tawk and C.A. contributed most of the experimental data. D.A.L., G.C.G and P.R.B. contributed additional experimental data. A.M.R. provided the *pard3–GFP* and pard3-Δ6–GFP constructs. D.R.H. provided Pard3 antisera and initial *Pard3* morpholino. M. Tada provided constructs and helped design experiments. J.D.W.C conceived the project, designed experiments and wrote the manuscript together with M.Tawk.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.D.W.C. (jonathan.clarke@ucl.ac.uk).

LETTERS

Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry

Matthew J. Evans[1]*, Thomas von Hahn[1]*, Donna M. Tscherne[1], Andrew J. Syder[1], Maryline Panis[1], Benno Wölk[1], Theodora Hatziioannou[2], Jane A. McKeating[1]†, Paul D. Bieniasz[2] & Charles M. Rice[1]

Hepatitis C virus (HCV) is a leading cause of cirrhosis and liver cancer worldwide. A better understanding of the viral life cycle, including the mechanisms of entry into host cells, is needed to identify novel therapeutic targets. Although HCV entry requires the CD81 co-receptor, and other host molecules have been implicated, at least one factor critical to this process remains unknown (reviewed in refs 1–3). Using an iterative expression cloning approach we identified claudin-1 (CLDN1), a tight junction component that is highly expressed in the liver[4], as essential for HCV entry. CLDN1 is required for HCV infection of human hepatoma cell lines and is the first factor to confer susceptibility to HCV when ectopically expressed in non-hepatic cells. Discrete residues within the first extracellular loop (EL1) of CLDN1, but not protein interaction motifs in intracellular domains, are critical for HCV entry. Moreover, antibodies directed against an epitope inserted in the CLDN1 EL1 block HCV infection. The kinetics of this inhibition indicate that CLDN1 acts late in the entry process, after virus binding and interaction with the HCV co-receptor CD81. With CLDN1 we have identified a novel key factor for HCV entry and a new target for antiviral drug development.

Virus entry often requires sequential interactions between viral proteins and multiple cellular factors. HCV virus particles are believed to consist of a nucleocapsid surrounded by a lipid bilayer studded with complexes of two envelope glycoproteins (HCVgp), E1 and E2. HCV entry is clathrin-dependent, requires a low pH compartment and occurs in a temperature-dependent manner[5–10]. Strong evidence exists for the involvement of glycosaminoglycans and two E2-binding proteins—scavenger receptor class B member I (SR-BI) and the tetraspanin CD81—in HCV entry (reviewed in refs 1–3). However, these molecules are insufficient for productive viral entry because some cell lines express all three factors but do not support HCV entry.

We performed a cyclic lentivirus based repackaging screen of a complementary DNA library, derived from the highly HCV-permissive hepatocarcinoma Huh-7.5 cell line[11], for genes that render the non-permissive CD81$^+$ SR-BI$^+$ 293T cell line infectable with HIV-1 particles pseudotyped with HCVgp (HCVpp) (Supplementary Fig. 1a and Supplementary Methods). As a result of this screen CLDN1 was identified as a potential HCV entry factor (Supplementary Fig. 1a, b). CLDN1 is one of 24 known claudin family members that form the backbone of tight junctions through homo- and heterotypic interactions (for a review see ref. 12). It is most highly expressed in liver, but is also found in several other epithelial tissues[4]. Expression of CLDN1 specifically enhanced 293T susceptibility to diverse genotypes of HCVpp by more than 100-fold, as determined by measuring luciferase activity encoded by these pseudoparticles (Supplementary Fig. 1c and Fig. 1a). CLDN1 did not affect 293T susceptibility to pseudoparticles bearing either no envelope proteins (Env$^-$pp) or the unrelated vesicular stomatitis virus G protein (VSV-Gpp), which serve as negative and positive infection controls, respectively.

We next sought to determine if CLDN1 overexpression renders 293T cells susceptible to infection with authentic cell culture grown HCV (HCVcc)[13]. Whereas naive 293T cells showed no NS5A$^+$ foci, a marker of ongoing virus replication, numerous NS5A$^+$ foci were detected in cells expressing CLDN1 (Fig. 1b). The effective titre of HCVcc was approximately 1,000-fold lower on the CLDN1 expressing 293T cells as compared with Huh-7.5 cells, probably reflecting low permissiveness for initiation of HCV RNA replication (as previously demonstrated for 293 cells when entry is bypassed by RNA transfection[11,14]) rather than a defect in entry as HCVpp infectivity is similar for CLDN1-expressing 293T and Huh-7.5 cells (Fig. 1a).

To test whether CLDN1 expression correlates with permissiveness for HCV entry, we surveyed the HCVpp susceptibility (Fig. 2a) and CLDN1 expression level (Fig. 2b) for human cell lines pre- and post-transduction with CLDN1 expression vectors. The HCVpp-susceptible Huh-7.5 and Hep3B human hepatoma cell lines expressed endogenous CLDN1 and overexpression of CLDN1 in these cells did not enhance HCV entry. Two HCVpp-resistant non-hepatic cell lines, 293T and SW13, did not express detectable levels of endogenous CLDN1, but became highly susceptible to HCVpp on CLDN1 expression. The CD81-negative hepatoma HepG2 line, previously

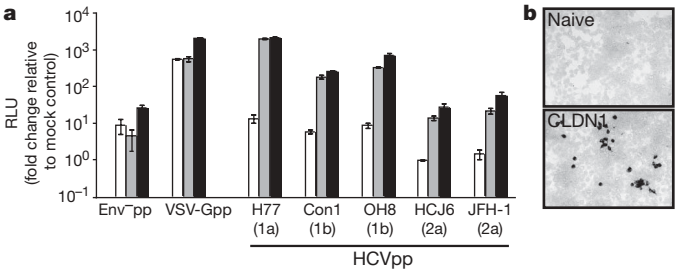


Figure 1 | CLDN1 expression confers susceptibility to HCV infection. **a**, Luciferase-encoding pseudoparticles bearing the indicated gps were used to infect naive (white) or CLDN1-expressing (grey) 293T cells, or Huh-7.5 cells (black). For HCVpp the isolate name is given and the respective genotype is indicated in parentheses. Values are normalized to the RLU background measured in mock-infected cells (mean of $n = 3$; error bars, s.d.). RLU, relative light units. **b**, Either naive or CLDN1-expressing 293T cells were infected with HCVcc (J6/JFH chimaeric genome). At 72 h post infection, cells were fixed with methanol and stained for NS5A antigen. Representative visual fields are shown.

[1]Center for the Study of Hepatitis C, The Rockefeller University, 1230 York Ave, New York 10021, USA. [2]Aaron Diamond AIDS Research Center, The Rockefeller University, 455 First Avenue, New York 10016, USA. †Present address: Division of Immunity and Infection, Institute of Biomedical Research, Medical School, University of Birmingham, Birmingham B15 2TT, UK.
*These authors contributed equally to this work.

reported to become susceptible on CD81 expression[6,15–17], also expressed endogenous CLDN1 and HCVpp entry remained CD81-dependent even when CLDN1 was overexpressed. However, the human HeLa and HepH cell lines (both $CD81^+$ SR-BI^+ $CLDN1^-$) remained HCVpp-resistant when overexpressing CLDN1 (Supplementary Fig. 6a). Likewise, expression of human CLDN1 and/or CD81 in non-human cell lines did not allow HCVpp entry (Supplementary Fig. 6b and data not shown). Moreover, murine CLDN1 (90% amino acid identity to human) efficiently supported HCV entry, indicating that CLDN1 is not a determinant of species host range (Supplementary Fig. 7). These observations suggest that one or more human-specific HCV entry factor(s) remain to be discovered.

To determine if endogenous CLDN1 is required for HCV infection of permissive liver cells, Hep3B cells were transfected with short interfering RNAs (siRNAs) targeting *CD81* or *CLDN1*. Expression of the respective proteins was determined by flow cytometry (Fig. 3a). Susceptibility to luciferase-encoding HCVpp (Fig. 3b), but not VSV-Gpp (Fig. 3b) was markedly reduced in *CLDN1*-silenced cells compared with cells treated with an irrelevant siRNA. When using six distinct siRNAs of variable potency against *CLDN1*, HCVpp infectivity was inversely correlated with the efficiency of silencing (Fig. 3c). Furthermore, CLDN1 knockdown in Huh-7.5 cells potently inhibited HCVcc infection (Fig. 3d, e). Given that even moderate downregulation of CLDN1 expression inhibited HCV infection, it is conceivable that the relatively low-level CLDN1 expression in extrahepatic tissues[4] may be insufficient to support HCV entry and thus CLDN1 may contribute to HCV tissue tropism.

Of the other claudin family members, neither CLDN7, the closest relative of CLDN1 (60% amino acid identity), nor CLDN3, the closest liver-expressed relative (49% identity), rendered 293T cells permissive to HCVpp (Fig. 4a), despite high levels of protein expression (Fig. 4b). To map CLDN1-specific entry determinants, we analysed a series of CLDN1–CLDN7 chimaeras (Fig. 4d) fused to the carboxy-terminus of GFP. The parental fusion proteins functioned identically to untagged proteins (Fig. 4e) and were expressed similarly (Fig. 4f). The expression and membrane-localization of all mutants were confirmed by flow cytometry and confocal microscopy, respectively (data not shown). CLDN1 is a 211 amino acid protein with four transmembrane helices, intracellular amino and carboxy termini and two extracellular loops (Fig. 4c)[4]. When extracellular loops were exchanged between CLDN1 and CLDN7 (Fig. 4d) it was found that only those containing the CLDN1 EL1 enhanced HCVgp-dependent infection (Fig. 4e). Progressively smaller exchanges identified the N-terminal third of the CLDN1 EL1 (EL1 N1/3) to be sufficient in an otherwise CLDN7 background to confer full susceptibility to HCVpp entry in 293T cells (Fig. 4e). The importance of CLDN1 EL1 in HCVpp entry was confirmed by deletion and insertional mutagenesis (Supplementary Fig. 2a–c).

Five residues differ between CLDN1 and CLDN7 in the critical region of EL1 (Fig. 4g). Although changing three of these in CLDN7 to the corresponding CLDN1 sequence had no effect, introduction of M32I or K48E into CLDN7 rendered 293T cells partially HCVpp permissive (46% and 14% of wild-type CLDN1, respectively;

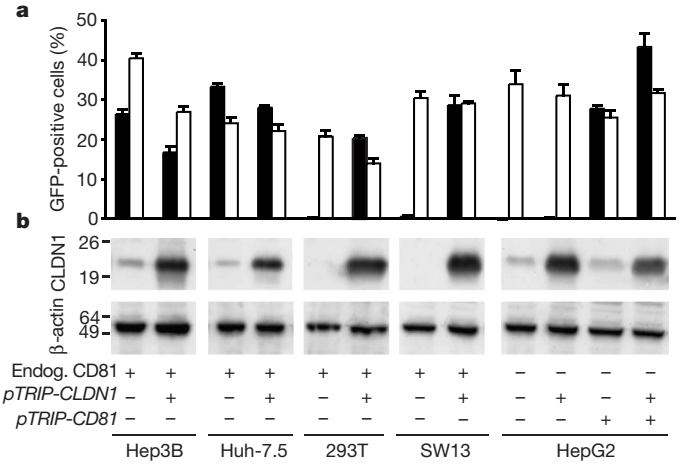


Figure 2 | CLDN1 expression is associated with susceptibility to HCVpp in human cell lines. **a**, The indicated human cell lines were either mock transduced or transduced to express CLDN1 ('*pTRIP-CLDN1*'), CD81 ('*pTRIP-CD81*'), or both and then challenged with HCVpp (black bars) and VSV-Gpp (white bars) encoding a GFP reporter. Background readings, that is, percentage of GFP-positive cells seen with Env^-pp, were subtracted for each population (mean of $n = 3$; error bars, s.d.). **b**, CLDN1 expression was assessed by immunoblotting. Approximate molecular weight (kDa) marker positions are indicated to the left of each blot. Cell surface CD81 expression ('endog. CD81') was determined before transduction, by flow cytometry using anti-CD81 1.3.3.22 as a primary antibody.

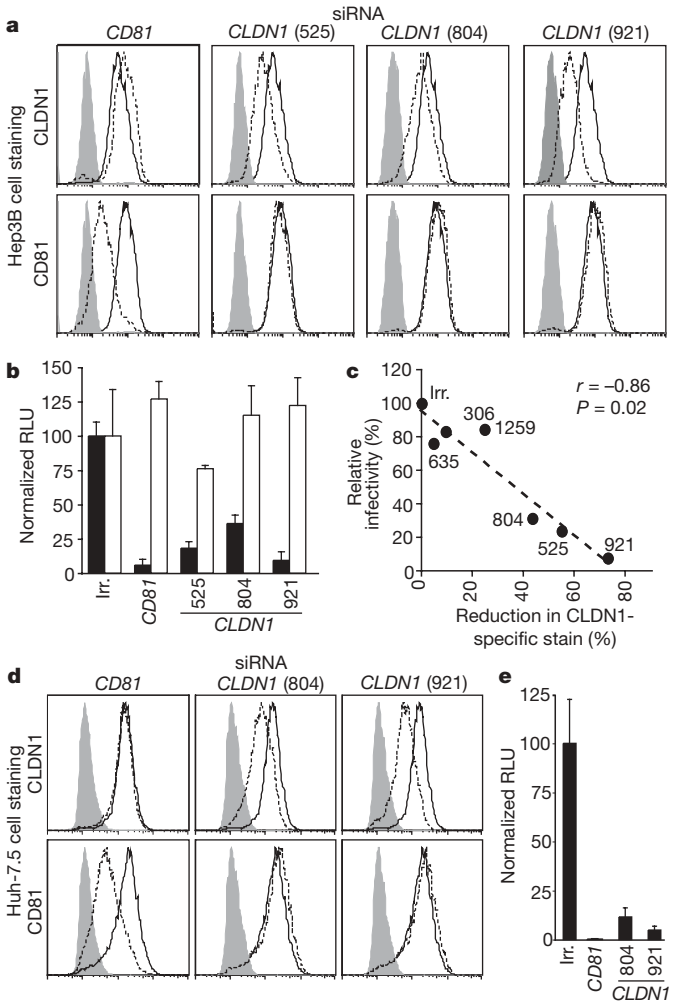


Figure 3 | CLDN1 silencing inhibits HCV entry. **a**, Dashed and solid lines represent Hep3B cells treated with specific and an irrelevant siRNA, respectively, and stained for CLDN1 (upper panels) or CD81 (lower panels). The shaded area represents isotype-control-staining of specific siRNA-treated cells. **b**, HCVpp (black bars) and VSV-Gpp (white bars) infection of siRNA-treated Hep3B cells normalized to irrelevant-siRNA-treated cells (irr) (mean of $n = 3$; error bars, s.d.). **c**, Correlation between CLDN1 downregulation and HCV infectivity using six distinct siRNAs against *CLDN1*. The *x* axis measures the reduction in Δ median fluorescence intensity (CLDN1 stain minus isotype control stain) relative to irrelevant-siRNA-treated cells; *y*-axis, HCVpp infectivity normalized to VSV-Gpp infectivity. *P*, probability for no correlation as determined by Spearman's non-parametric test. **d**, Silencing of CD81 and CLDN1 in Huh-7.5 cells. **e**, siRNA-treated Huh-7.5 cells were infected with HCVcc encoding a luciferase reporter (mean of $n = 4$; error bars, s.d.).

Fig. 4h) and the combination of both changes supported HCV entry as efficient as wild-type CLDN1. Conversely, introduction of I32M or E48K into CLDN1 dramatically reduced HCVpp susceptibility, and the presence of both mutations reduced entry to background levels (Fig. 4h). None of the wild-type or mutant claudin molecules affected VSV-Gpp entry (Fig. 4h).

These data suggest that any CLDN1 features required for HCV infection other than I32 and E48 must be conserved between CLDN1 and CLDN7. In additional analyses, we found the CLDN1 C-terminal intracellular tail, which mediates interactions with other TJ proteins[18], and putative palmitoylation sites, which may regulate protein interactions or direct the molecule to specific regions of the plasma membrane[12], were not required for HCVpp entry (Supplementary Fig. 3a, b). This suggests a direct involvement of CLDN1 in HCV entry through an extracellular interaction rather than an indirect effect mediated through intracellular interactions with other tight junction components. All claudin mutants were also tested in SW13 cells with similar results (data not shown), confirming that the observed phenotypes were not 293T cell specific.

Blocking antibodies, ligands or antagonists are useful for probing the function of cellular molecules involved in viral entry. Unfortunately, all available CLDN1-specific antibodies recognize the intracellular C-terminal segment of the protein and are thus not useful for such studies. After an unsuccessful attempt to raise antibodies against CLDN1 EL1 (data not shown), we identified a site in the C-terminal portion of CLDN1 EL1 where a triple Flag epitope sequence could be inserted without seriously impairing HCV entry (Supplementary Fig. 2a, c). HCVpp infection of 293T cells expressing this mutant (CLDN1 F10x3) was blocked in a dose-dependent manner by anti-Flag M2 monoclonal antibody. As controls, neither HCVpp infection of cells expressing wild-type CLDN1 nor VSV-Gpp infection of cells expressing CLDN1 F10x3 were affected by equal amounts of antibody (Fig. 5a and Supplementary Fig. 4). Although these results further suggest that CLDN1 functions in HCV entry through a direct interaction between EL1 and the virion, evidence for direct HCV binding to CLDN1 is lacking. In fact, HCVcc binding to CHO cells was only enhanced by expression of human SR-BI, but not CLDN1 or CD81 (Supplementary Fig. 8). However, virus–receptor interactions before CLDN1 engagement may trigger HCV glycoprotein conformational changes required for CLDN1 binding, paralleling the situation with HIV-1 and its co-receptor CCR-5, in which binding requires prior interaction with CD4 (ref. 19).

To determine when CLDN1 functions in HCV entry, we examined the ability of the M2 antibody to block infection of CLDN1 F10x3 cells when added at various times during cell entry. To synchronize infection, 293T cells expressing CLDN1 F10x3 were incubated with HCVpp for 2 h at 4 °C, so that virion binding but not entry could occur, and then washed and shifted to 37 °C, allowing the infection to continue (see Supplementary Methods). Antibodies directed against both CD81 (JS81) and the Flag epitope (M2) inhibited HCVpp infection and retained maximal inhibitory activity even after the temperature shift to 37 °C, indicating that like CD81 (ref. 10), CLDN1 acts at the post-binding stage of HCV entry (Fig. 5b). However, the inhibitory activity of anti-CD81 was lost much earlier than that of anti-Flag (half-maximal inhibition at 18 and 73 min post temperature shift, respectively), suggesting a sequence of events in which CD81 acts prior to CLDN1 in HCV entry.

To test if CLDN1 is required for HCVgp-mediated membrane fusion we used a cell–cell fusion assay where 293T 'acceptor' cells, encoding a HIV-1 transactivator of transcription (Tat)-dependent GFP, were co-cultured with Tat-expressing 'donor' cells such that fusion between donor and acceptor cell membranes results in GFP expression (Supplementary Fig. 5 and Supplementary Methods). When using donor cells expressing HCV E1E2, expression of CLDN1 on the acceptor cells resulted in a sevenfold increase in the number of GFP-positive foci per well ($P < 0.001$ by unpaired t-test), indicating a significant enhancement of HCVgp-dependent cell–cell fusion (Fig. 5c). A pH 5 wash did not significantly increase the number of GFP foci observed, but changed their appearance from clusters of discrete GFP-positive cells to large brightly GFP-positive syncytia, indicating that the pH-dependence of HCVgp-mediated fusion[5–10] is maintained in this assay (Supplementary Fig. 5). These results demonstrate that CLDN1 is required for HCVgp-dependent cell fusion although it is presently unclear whether the protein participates directly in the fusion process or acts at an earlier step that is required to enable subsequent fusion.

The requirement for numerous cellular factors including glycosaminoglycans, SR-BI, CD81 and now CLDN1, indicates that HCV cell entry is a complex multi-step process (reviewed in refs 1–3). An

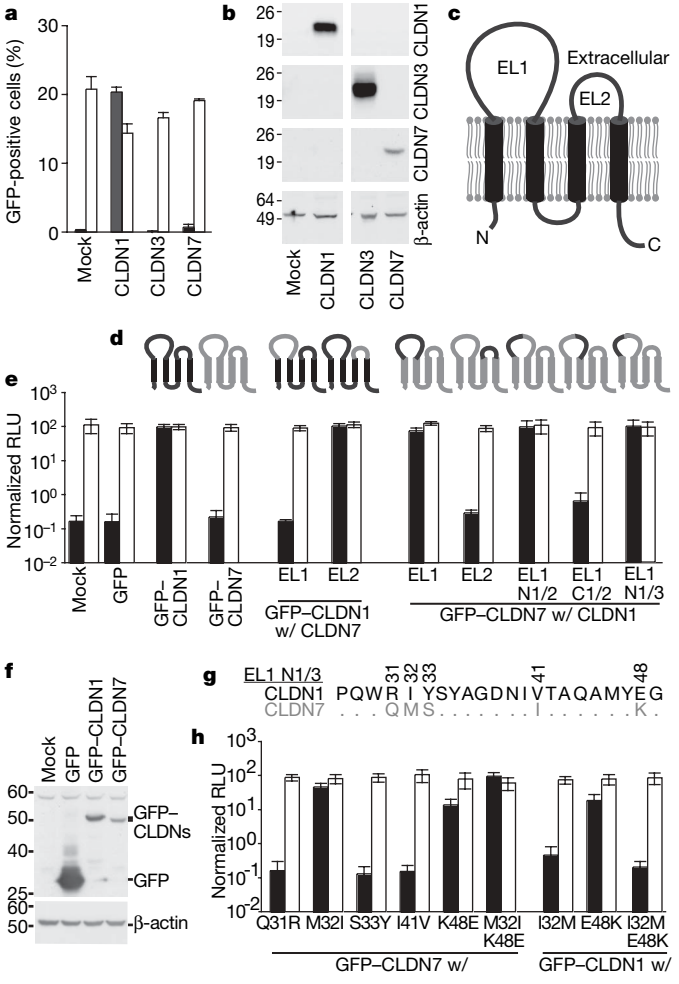


Figure 4 | HCVpp susceptibility depends on residues in the first extracellular loop of CLDN1. **a**, GFP-reporter HCVpp (black) and VSV-Gpp (white) infection of 293T cells expressing CLDN1, 3 or 7. Background readings, that is, percentage of GFP-positive cells seen with Env$^-$pp, were subtracted for each population (mean of $n = 3$, error bars, s.d.). **b**, Immunoblot for CLDN1, 3 and 7. Approximate molecular weight (kDa) marker positions are indicated to the left of each blot. **c**, CLDN1 topology. **d**, CLDN1/CLDN7 chimaeras with regions of CLDN1 and CLDN7 represented as dark and light lines, respectively. All chimaeras were N-terminally GFP-tagged. **e**, HCVpp (black) and VSV-Gpp (white) infection of 293T cells expressing the chimaeras depicted above in panel **d**. For chimeras, the x axis labels refer to the GFP–CLDN1 or GFP–CLDN7 fusion backbone encoding ('w/') the indicated region of the other claudin protein. N and C indicate either the amino or carboxy terminal, respectively, piece of CLDN1 EL1, swapped into CLDN7. Readings are normalized to 293T cells expressing wild-type CLDN1 (mean of $n = 4$; error bars, s.d.). **f**, Immunoblot for GFP. **g**, Alignment of the N-terminal half of EL1 in CLDN1 and CLDN7. Identical sequences are represented by a full-stop, numbering represents amino acid position in full-length CLDN1. **h**, 293T cells expressing the indicated point mutants were tested for susceptibility as described above (mean of $n = 4$; error bars, s.d.).

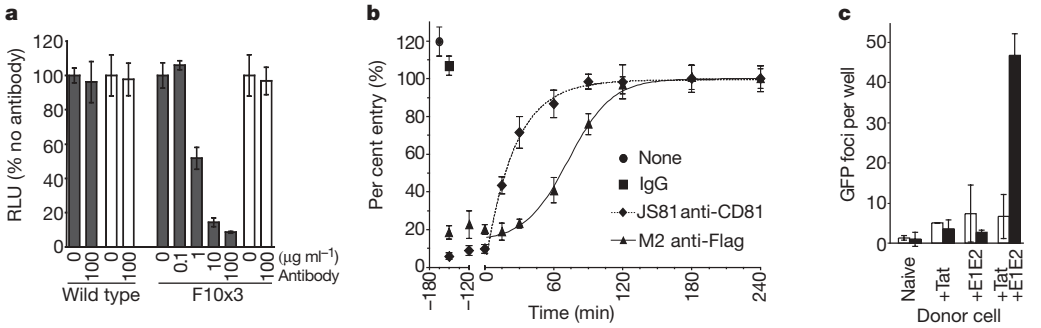


Figure 5 | Analysis of CLDN1 function in the HCV entry process. a, 293T cells expressing wild-type or F10x3 CLDN1 were challenged with HCVpp (black) or VSV-Gpp (white) in the presence or absence of M2 anti-Flag. (Values were normalized to RLU from cells with no antibody; mean of $n = 4$; error bars, s.d.). **b**, Synchronized infections were performed on 293T cells expressing F10x3 CLDN1, with the indicated antibodies being present from the time indicated on the *x* axis. Values (per cent entry) are relative to the signal seen when antibody was added 4 h post temperature shift. Controls are normalized to the value for M2 added at 4 h. Fits of $t = 0$ and later data points represent a one-phase exponential association and sigmoidal dose–response (variable slope) for JS81 and M2, respectively (mean of $n = 3$; error bars, s.d.). **c**, Cell fusion assay using 293T acceptor cells expressing Tat-regulated GFP alone (white) or together with CLDN1 (black), in co-culture with 293T donor cells expressing no transgene (naive), HIV-1 Tat, HCV E1E2 or both (mean of $n = 3$; error bars, s.d.).

important challenge is to delineate the precise function that each cellular entry factor fulfils. Our data suggest a role for CLDN1 late in the HCV entry process, perhaps downstream of interactions with other virus-receptor/co-receptors such as CD81. However, differences in blocking antibody efficacy are an important caveat to bear in mind when interpreting the inhibition data in Fig. 5b.

Preliminary studies indicate that CLDN1 is localized on the plasma membrane in HCV permissive cells, is enriched at sites of cell–cell contact and co-localizes with a sub-fraction of CD81 (unpublished data). Interestingly, an association of tetraspanins with claudin-11 has been reported[20]. It should be emphasized, however, that the situation may differ in polarized hepatocytes, where CLDN1 is strictly localized to tight junction regions[21]. This location fits well with a co-receptor rather than a primary receptor role for CLDN1. Other tight-junction-associated molecules have been found to serve as (co-) receptors for reoviruses[22], coxsackievirus B and certain adenoviruses[23]. Coxsackievirus B entry requires initial binding to a primary receptor (decay-accelerating factor) on the luminal cell surface, followed by lateral migration of the virus–receptor complex to the tight junction, where interaction with the CAR co-receptor and uptake into the host cell occur[24]. A similar pathway is conceivable for HCV, and it will be of considerable interest to devise systems to examine HCV entry in polarized cells.

The involvement of CLDN1 in HCV entry raises interesting questions regarding the pathogenesis of hepatitis C. Loss of cell–cell contacts is a hallmark of malignant transformation and altered expression of CLDN1 and other claudin family members has been reported in a number of human malignancies including hepatocellular carcinoma—the most lethal complication of chronic HCV infection[25–27]. In addition, the extremely high viral loads observed in some immunosuppressed liver transplant recipients could conceivably compromise tight junction integrity and contribute to the development of cholestatic hepatitis[28,29]. Finally, CLDN1 presents a new target for HCV therapeutic intervention that may complement ongoing efforts to block intracellular replication events with inhibitors of the HCV proteases and polymerase.

METHODS

For more detailed methods see Supplementary Information.

Pseudoparticles. All pseudoparticles were generated by 293T FuGENE6 (Roche Applied Science, Indianapolis) cotransfection of plasmids encoding (1) a minimal HIV provirus (such as pTRIP, V1 or CSxW) encoding a reporter gene or other transgene, (2) HIV gag-pol and (3) an appropriate viral gps (HCV E1E2 or VSV-G). The pNL43.luc.R^-.E^- provirus encodes gag-pol and only requires the addition of the gps[5]. To generate Env^-pp the glycoprotein vector was replaced with empty vector. Pseudoparticle-containing supernatants were harvested at 48 and 72 h were pooled and filtered (0.45 μm mesh). Pseudoparticle infections were performed in the presence of 4 μg ml^{-1} polybrene. A minimum of 48 h elapsed between transduction and reporter gene quantification or subsequent experiments.

Cyclic lentivirus cDNA library repackaging screen. The V1 lentiviral Huh-7.5-derived cDNA library was packaged into VSV-Gpp and applied to 293T cells, which were then challenged with HCVpp, encoding either a puromycin (CSPW) or zeocin (CSZW) reporter gene, and subjected to antibiotic selection (Supplementary Fig. 1a). Surviving clones were pooled and a fraction tested for their susceptibility to GFP-encoding HCVpp. The populations were subsequently either challenged with a different antibiotic resistance gene for further selection or transfected with HIV gag-pol and VSV-G expression plasmids to package the V1 genomes with the cDNA inserts contained in the surviving 293T cells into fresh VSV-Gpp (cyclic packaging rescue, CPR). These were used to transduce naive 293T cells for subsequent rounds of screening. Deletions in the 3′-long terminal repeat of the CSPW and CSZW proviruses prevented their re-packaging such that the newly transduced 293T cells could again be challenged as outlined above.

Cell culture grown HCV (HCVcc). Plasmids encoding the chimaeric J6/JFH genome[13] and FL-J6/JFH-5′C19Rluc2AUbi reporter genome[9] were *Xba*I linearized and transcribed using MEGAscript T7 (Ambion, Austin, Texas). RNA was electroporated into Huh-7.5 cells using a ECM 830 electroporator (BTX Genetronics). High-titre stocks were generated by serial passage through naive Huh-7.5 cells.

Received 18 December 2006; accepted 6 February 2007.
Published online 25 February; corrected 12 April 2007 (details online).

1. Bartosch, B. & Cosset, F. L. Cell entry of hepatitis C virus. *Virology* **348**, 1–12 (2006).
2. Cocquerel, L., Voisset, C. & Dubuisson, J. Hepatitis C virus entry: potential receptors and their biological functions. *J. Gen. Virol.* **87**, 1075–1084 (2006).
3. Barth, H., Liang, T. J. & Baumert, T. F. Hepatitis C virus entry: molecular biology and clinical implications. *Hepatology* **44**, 527–535 (2006).
4. Furuse, M., Fujita, K., Hiiragi, T., Fujimoto, K. & Tsukita, S. Claudin-1 and -2: novel integral membrane proteins localizing at tight junctions with no sequence similarity to occludin. *J. Cell Biol.* **141**, 1539–1550 (1998).
5. Hsu, M. *et al.* Hepatitis C virus glycoproteins mediate pH-dependent cell entry of pseudotyped retroviral particles. *Proc. Natl Acad. Sci. USA* **100**, 7271–7276 (2003).
6. Bartosch, B. *et al.* Cell entry of hepatitis C virus requires a set of co-receptors that include the CD81 tetraspanin and the SR-B1 scavenger receptor. *J. Biol. Chem.* **278**, 41624–41630 (2003).
7. Meertens, L., Bertaux, C. & Dragic, T. Hepatitis C Virus entry requires a critical postinternalization step and delivery to early endosomes via clathrin-coated vesicles. *J. Virol.* **80**, 11571–11578 (2006).
8. Blanchard, E. *et al.* Hepatitis C virus entry depends on clathrin-mediated endocytosis. *J. Virol.* **80**, 6964–6972 (2006).
9. Tscherne, D. M. *et al.* Time- and temperature-dependent activation of hepatitis C virus for low-pH-triggered entry. *J. Virol.* **80**, 1734–1741 (2006).
10. Koutsoudakis, G. *et al.* Characterization of the early steps of hepatitis C virus infection by using luciferase reporter viruses. *J. Virol.* **80**, 5308–5320 (2006).
11. Blight, K. J., McKeating, J. A. & Rice, C. M. Highly permissive cell lines for subgenomic and genomic hepatitis C virus RNA replication. *J. Virol.* **76**, 13001–13014 (2002).

12. Van Itallie, C. M. & Anderson, J. M. Claudins and epithelial paracellular transport. *Annu. Rev. Physiol.* **68,** 403–429 (2006).
13. Lindenbach, B. D. *et al.* Complete replication of hepatitis C virus in cell culture. *Science* **309,** 623–626 (2005).
14. Kato, T. *et al.* Nonhepatic cell lines HeLa and 293 support efficient replication of the hepatitis C virus genotype 2a subgenomic replicon. *J. Virol.* **79,** 592–596 (2005).
15. Zhang, J. *et al.* CD81 is required for hepatitis C virus glycoprotein-mediated viral infection. *J. Virol.* **78,** 1448–1455 (2004).
16. McKeating, J. A. *et al.* Diverse hepatitis C virus glycoproteins mediate viral infection in a CD81-dependent manner. *J. Virol.* **78,** 8496–8505 (2004).
17. Cormier, E. G. *et al.* CD81 is an entry coreceptor for hepatitis C virus. *Proc. Natl Acad. Sci. USA* **101,** 7270–7274 (2004).
18. Heiskala, M., Peterson, P. A. & Yang, Y. The roles of claudin superfamily proteins in paracellular transport. *Traffic* **2,** 93–98 (2001).
19. Wu, L. *et al.* CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature* **384,** 179–183 (1996).
20. Tiwari-Woodruff, S. K. *et al.* OSP/claudin-11 forms a complex with a novel member of the tetraspanin super family and β1 integrin and regulates proliferation and migration of oligodendrocytes. *J. Cell Biol.* **153,** 295–305 (2001).
21. Furuse, M. *et al.* Claudin-based tight junctions are crucial for the mammalian epidermal barrier: a lesson from claudin-1-deficient mice. *J. Cell Biol.* **156,** 1099–1111 (2002).
22. Barton, E. S. *et al.* Junction adhesion molecule is a receptor for reovirus. *Cell* **104,** 441–451 (2001).
23. Bergelson, J. M. *et al.* Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. *Science* **275,** 1320–1323 (1997).
24. Coyne, C. B. & Bergelson, J. M. Virus-induced Abl and Fyn kinase signals permit coxsackievirus entry through epithelial tight junctions. *Cell* **124,** 119–131 (2006).
25. Borlak, J., Meier, T., Halter, R., Spanel, R. & Spanel-Borowski, K. Epidermal growth factor-induced hepatocellular carcinoma: gene expression profiles in precursor lesions, early stage and solitary tumours. *Oncogene* **24,** 1809–1819 (2005).
26. Swisshelm, K., Macek, R. & Kubbies, M. Role of claudins in tumorigenesis. *Adv. Drug Deliv. Rev.* **57,** 919–928 (2005).
27. Dhawan, P. *et al.* Claudin-1 regulates cellular transformation and metastatic behavior in colon cancer. *J. Clin. Invest.* **115,** 1765–1776 (2005).
28. Doughty, A. L., Spencer, J. D., Cossart, Y. E. & McCaughan, G. W. Cholestatic hepatitis after liver transplantation is associated with persistently high serum hepatitis C virus RNA levels. *Liver Transpl. Surg.* **4,** 15–21 (1998).
29. Schluger, L. K. *et al.* Severe recurrent cholestatic hepatitis C following orthotopic liver transplantation. *Hepatology* **23,** 971–976 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank J. Tassello, M. Hunter and N. Torres for excellent technical assistance; S. You for providing HCVcc/Rluc virus stocks; D. Schmid and M. Landthaler for expert advice on RNAi; and M. MacDonald and L. Dustin for reviewing the manuscript. This work was supported by the Greenberg Medical Research Institute, the Ellison Medical Foundation, the Starr Foundation, the Ronald A. Shellow Memorial Fund, the Richard Salomon Family Foundation, and the National Institutes of Health (grants to M.J.E., T.v.H., D.M.T., A.J.S., P.D.B. and C.M.R.). P.D.B. is an Elizabeth Glaser Pediatric AIDS Foundation Scientist. C.M.R. is an Ellison Medical Foundation Senior Scholar in Global Infectious Diseases. T.v.H. and B.W. were supported by postdoctoral fellowships from the Deutsche Forschungsgemeinschaft. This work was presented in part at the 13th International Meeting on Hepatitis C Virus & Related Viruses, Cairns, Australia, 27–31 August, 2006.

Author Contributions M.E., T.v.H. and C.M.R. designed the project, analysed results and wrote the manuscript. M.E., T.v.H., D.M.T., A.J.S., M.P. and B.W. performed the experimental work. T.H. and P.D.B developed the screening technology and assisted in its implementation. T.v.H. and J.A.M were involved in preliminary experiments identifying and characterizing HCV nonpermissive cell lines.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to C.M.R. (ricec@mail.rockefeller.edu).

LETTERS

Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map

Sean R. Collins[1,2,3], Kyle M. Miller[4], Nancy L. Maas[4], Assen Roguev[1,2], Jeffrey Fillingham[5], Clement S. Chu[1,2,3], Maya Schuldiner[1,2,3], Marinella Gebbia[5], Judith Recht[6], Michael Shales[5], Huiming Ding[5], Hong Xu[5], Junhong Han[7], Kristin Ingvarsdottir[8], Benjamin Cheng[9], Brenda Andrews[5], Charles Boone[5], Shelley L. Berger[8], Phil Hieter[9], Zhiguo Zhang[7], Grant W. Brown[10], C. James Ingles[5], Andrew Emili[5], C. David Allis[6], David P. Toczyski[4], Jonathan S. Weissman[1,2,3], Jack F. Greenblatt[5] & Nevan J. Krogan[1,2]

Defining the functional relationships between proteins is critical for understanding virtually all aspects of cell biology. Large-scale identification of protein complexes has provided one important step towards this goal; however, even knowledge of the stoichiometry, affinity and lifetime of every protein–protein interaction would not reveal the functional relationships between and within such complexes. Genetic interactions can provide functional information that is largely invisible to protein–protein interaction data sets. Here we present an epistatic miniarray profile (E-MAP)[1] consisting of quantitative pairwise measurements of the genetic interactions between 743 *Saccharomyces cerevisiae* genes involved in various aspects of chromosome biology (including DNA replication/repair, chromatid segregation and transcriptional regulation). This E-MAP reveals that physical interactions fall into two well-represented classes distinguished by whether or not the individual proteins act coherently to carry out a common function. Thus, genetic interaction data make it possible to dissect functionally multi-protein complexes, including Mediator, and to organize distinct protein complexes into pathways. In one pathway defined here, we show that Rtt109 is the founding member of a novel class of histone acetyltransferases responsible for Asf1-dependent acetylation of histone H3 on lysine 56. This modification, in turn, enables a ubiquitin ligase complex containing the cullin Rtt101 to ensure genomic integrity during DNA replication.

The synthetic genetic array (SGA)[2] and diploid-based synthetic lethality analysis on microarray (dSLAM)[3] approaches have enabled systematic identification of synthetic sickness/lethal (SSL) relationships in *S. cerevisiae* in which pairs of gene deletions are far more deleterious together than either of the individual deletions. Although individual SSL interactions can be difficult to interpret, the patterns of genetic interactions for gene mutations can be more informative because they provide high-resolution phenotypes that can be compared to identify functionally related genes[1–5]. Recently, we exploited the SGA strategy for generating double mutants to develop an approach, termed E-MAP[1], that greatly facilitates such comparisons. An E-MAP comprises comprehensive and quantitative measurements of genetic interactions between pairs of mutations within a defined subset of genes linked to one or more specific biological processes[1]. E-MAPs are created by systematically generating yeast strains carrying each pair of mutations and measuring their growth rates. Genetic interactions are determined by comparing the observed fitness of the double mutants to an empirically determined typical fitness that would be expected on the basis of the growth defects associated with each mutation[1,6]. This technique allows for the identification of not only negative (aggravating) interactions, such as SSL pairs, but also positive (alleviating) interactions. Positive interactions include suppression, in which double mutants are healthier than the sicker of the two single mutants, as well as cases in which loss of one gene masks the effect of losing another, as is seen when two proteins act together in a common complex or pathway.

We comprehensively evaluated pairwise genetic interactions for 754 alleles of 743 genes involved in various aspects of chromosome biology (Fig. 1a; see also Supplementary Fig. 1 and Supplementary Data). The mutations include deletions of 663 non-essential genes and constitutive hypomorphic alleles—constructed using the 'decreased abundance by messenger RNA perturbation' (DAmP) strategy[1]—for 70 essential genes. Genes were selected based on published functional studies, protein–protein interaction data[7,8], earlier genome-wide SSL studies[2] and chemical sensitivity screens.

The resulting E-MAP consists of a 754 by 754 matrix of genetic interaction scores, where each row corresponds to the pattern (or profile) of interactions for one mutant allele of a gene (Fig. 1a). Using hierarchical clustering, we reordered the matrix to sort genes according to the similarity of their genetic interaction profiles. The resulting map has a modular structure that distinguishes between major processes such as transcription and chromatin remodelling, DNA replication and repair, and sister chromatid segregation. We illustrate the high-resolution functional information within these modules by focusing on a subcluster containing genes involved in DNA replication and repair (Fig. 1b). The general DNA replication factors (for example, RPA (*RFA1* and *RFA2*) and RFC processivity clamp loader subunits (*RFC4* and *RFC5*)) cluster near each other, and are resolved from the DNA replication checkpoint complex, Mrc1–Csm3–Tof1. The E-MAP also distinguishes groups of genes involved in sensing and repairing DNA damage including the *RAD52* epistasis group (*RAD51*, *RAD52*, *RAD54*, *RAD55*, *RAD57*), the MRX complex

[1]Department of Cellular and Molecular Pharmacology, [2]The California Institute for Quantitative Biomedical Research, and [3]Howard Hughes Medical Institute, University of California, San Francisco, California 94158, USA. [4]Department of Biochemistry and Biophysics, Cancer Research Institute, University of California, San Francisco, California 94115, USA. [5]Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada. [6]Laboratory of Chromatin Biology, The Rockefeller University, New York, New York 10021, USA. [7]Department of Biochemistry and Molecular Biology, Mayo Clinic, College of Medicine, Rochester, Minnesota 55905, USA. [8]Gene Expression and Regulation Program, The Wistar Institute, Philadelphia, Pennsylvania 19104, USA. [9]Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. [10]Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada.

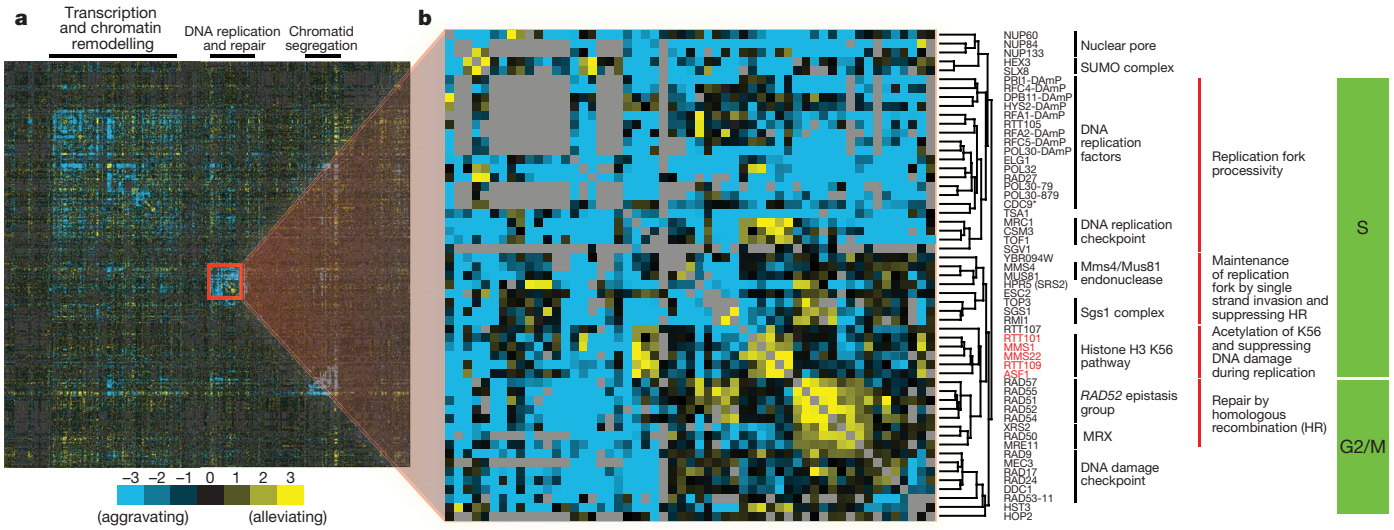


Figure 1 | Hierarchical clustering of genetic interaction patterns. **a**, Full 'clustergram' of the patterns of interactions for all 754 mutations. Black horizontal bars indicate regions of the cluster corresponding to sets of genes implicated in the indicated functional processes. **b**, An enlargement of the 'DNA replication and repair' subcluster from **a**. The dendrograms indicate the relative similarities of the full patterns of interactions for the indicated genes. Various subclusters are annotated according to their known functions. Blue and yellow represent negative and positive genetic interactions, respectively. Grey boxes correspond to missing data points. Here, and throughout, genes indicated with an asterisk correspond to deletions of spurious ORFs that overlap the indicated gene. Genes highlighted in red represent novel findings that are referred to in the text.

(*RAD50*, *MRE11*, *XRS2*) and the 9-1-1 clamp (*DDC1*, *MEC3*, *RAD17*). The complete genetic interaction map, a useful resource for future functional studies, is available in Supplementary Data.

Beyond allowing simple hierarchical clustering, patterns of genetic interactions provide an unbiased way to identify sets of genes that function together in a coherent manner[1–3,5]. If two proteins act together to carry out a common function, one would expect deletions of the two encoding genes to have highly similar profiles of genetic interactions, as deletion of either gene would disrupt the same cellular process. Similarly, one would expect a positive genetic interaction between the two deletions, because in the context of the first, the second deletion would incur no additional cost. This relationship can be formalized in a COP (complex or pathway) score (see Supplementary Methods)[1,6], which synthesizes both expectations to create a single mathematical metric. Sets of genes connected by high COP scores are analogous to classically defined epistasis groups such as the well-studied *RAD52* epistasis group (Fig. 1b and Supplementary Fig. 2).

Recently, we used large-scale affinity purification data[7,8] to generate a physical interaction map that quantitatively reports through a purification enrichment (PE) score on the relative likelihood of each protein–protein interaction (see http://interactome-cmp.ucsf.edu)[9].

The accuracy and completeness of this integrated physical interaction map and the present E-MAP now make it possible to explore broadly the relationship between physical complexes and genetically defined epistasis groups. To evaluate the predictive power of the COP score relative to the physically based PE score, we used the protein complexes in the *Saccharomyces* Genome Database (SGD)[10] to define a trusted reference set of 'true positives' and 'true negatives' (see Supplementary Methods). We then used receiver operating characteristic (ROC) curves[11], which measure the rate at which each approach identifies true positives versus true negatives, to compare the predictive power of the two approaches.

Notably, the COP score identifies a distinct and large subset of protein–protein interactions with a specificity rivalling that of affinity purification (Fig. 2a; see also Supplementary Fig. 2). A key value of the E-MAP, therefore, is that it divides physical interactions into two classes: one group in which the proteins function coherently and a second in which their patterns of genetic interactions indicate that the proteins carry out distinct or even opposing functions. In particular, for pairs of physically interacting proteins, the histogram of either the genetic interaction scores or the correlation between genetic interaction patterns shows a roughly bimodal character (Fig. 2b, c). Thus, for a large fraction (somewhat greater than half) of physical

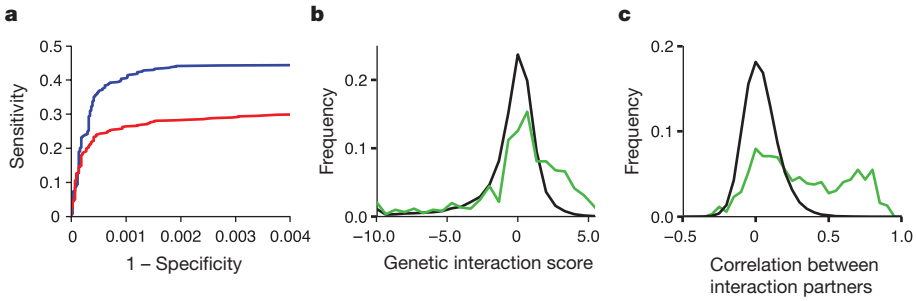


Figure 2 | Relationship between genetic epistasis groups and physical complexes. **a**, ROC curves comparing the power of the genetic interaction patterns—using the COP score (red) (see Supplementary Methods)—and large-scale affinity purification data (blue)—using a recent re-analysis of raw purification data[9]—to predict co-membership of pairs of proteins in the same physical complex. The slope of the initial portion of each curve serves as a measure of the score's maximal accuracy. **b**, Distribution of direct genetic interaction scores for pairs of genes encoding physically interacting proteins (green) and non-interacting proteins (black) (see Supplementary Data). **c**, Distribution of the Pearson's correlation coefficients between the interaction patterns for the same sets of gene pairs as in **b**.

interactions, individual deletions of the two genes cause largely unrelated phenotypes (see Supplementary Data for a complete comparison between genetic and physical complexes).

This ability of genetic interaction maps to reveal whether physically interacting proteins mediate a common function allows for the systematic dissection of multi-protein complexes into functional modules. We illustrate this by analysing Mediator, a conserved 25-protein complex critical for regulation of transcriptional initiation[12] by RNA polymerase II. Previous studies have suggested that Mediator is composed of four discrete physical modules: the head, middle, tail and CDK modules (Fig. 3b)[13,14]. Genetic interaction patterns accurately distinguish the protein components of these modules (Fig. 3a) and reveal distinct patterns of genetic interactions between components of different modules (Fig. 3a). For example, the CDK module displays positive genetic interactions with the head, middle and tail modules (Fig. 3a, b), consistent with the model in which the CDK module opposes the action of the remainder of the complex[14].

Our E-MAP data also reveal close functional relationships between Mediator components and non-Mediator proteins (Fig. 3a), thereby providing insights into specialized roles of the Mediator modules. For example, the transcriptional elongation factor TFIIS (Dst1) has a pattern of interactions highly correlated with patterns of middle module components, suggesting that this module may function with Dst1 during or after promoter escape. Conversely, the activator Swi4 clusters with the tail components, consistent with previous evidence that transcriptional activator proteins sometimes recruit Mediator through physical interactions with the tail module[13]. Finally, mutations of head module components behave most like mutations in the core transcriptional machinery (RNA polymerase II (*RPB3*, *IWR1*), TFIIF (*TFG1*, *TFG2*) and TFIIH (*RAD3*)), indicating that this module may have a critical role in the assembly and/or activity of the pre-initiation complex, which is consistent with a recent biochemical analysis[15].

In addition to dividing large protein complexes into functional sub-modules, the E-MAP-derived epistasis groups reveal many cases where proteins cooperate with each other even when they do not physically interact. One prominent and previously uncharacterized example of such an epistasis group is found within the DNA damage cluster (Fig. 1b) and comprises five genes: *RTT101*, *MMS1*, *MMS22*, *RTT109* and *ASF1* (Fig. 4a). Rtt101 is one of four cullins in budding yeast and is functionally[3] and physically associated with Mms1, Mms22, the E2 ubiquitin conjugating enzyme Cdc34, and the RING finger protein Hrt1 (refs 8–10, 16). This complex thus has features characteristic of ubiquitin ligases that target proteins for proteasomal degradation. *RTT109* is a poorly characterized open reading frame (ORF) identified as a regulator of Ty1 transposition[10]. Finally, Asf1 is a well-studied histone chaperone implicated in several

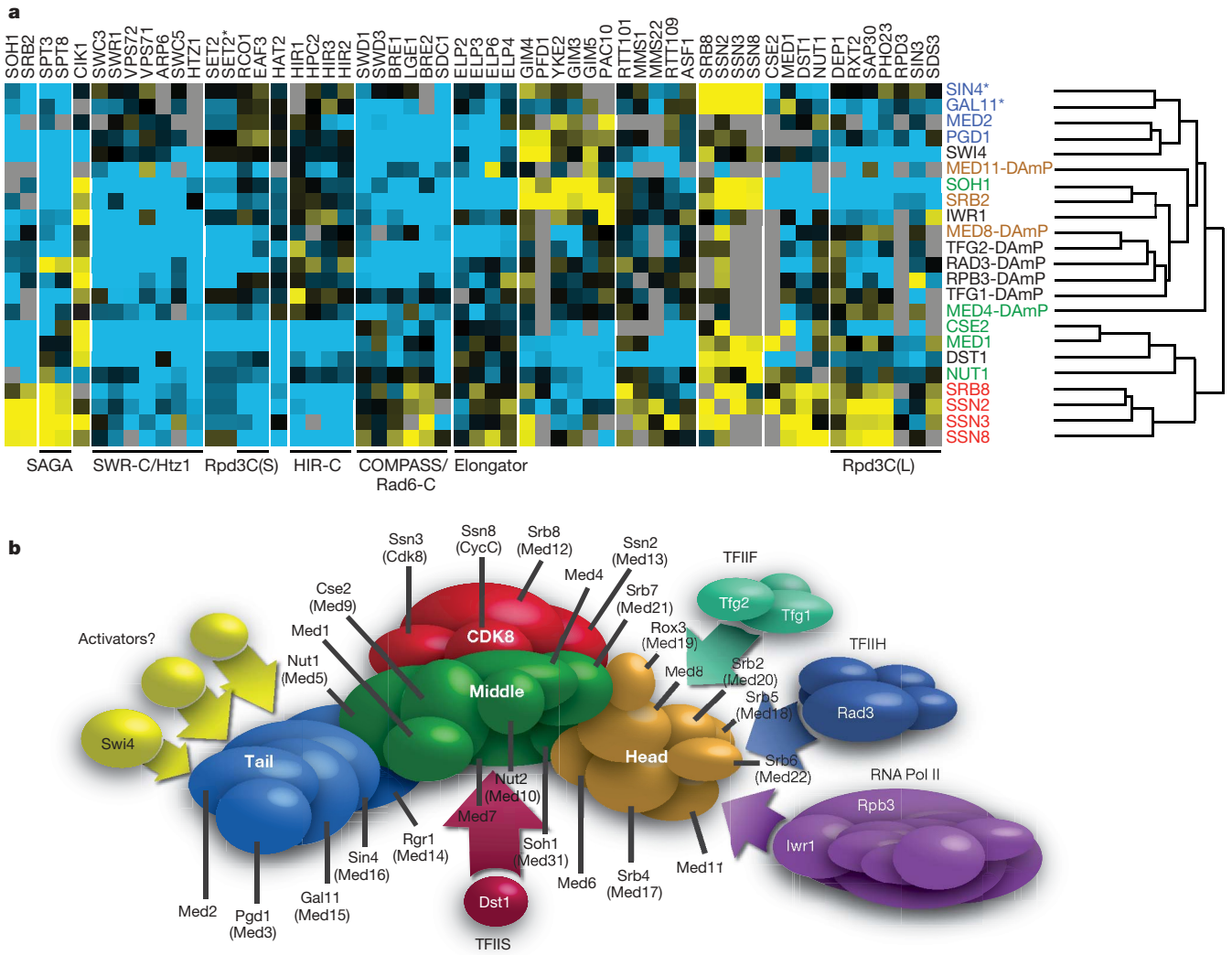


Figure 3 | Functional dissection of the Mediator complex. a, Subsets of the genetic interaction profiles for deletion or DAmP alleles of the indicated Mediator components (red, blue, green or orange font corresponding to the CDK, tail, middle and head modules, respectively) or non-Mediator genes (black type). The dendrogram indicates the relative similarities of the full interaction patterns for the mutations shown. Blue and yellow represent negative and positive genetic interactions, respectively. Grey boxes correspond to missing data points. **b**, Modular organization of the Mediator complex[14] and close functional relationships between the Mediator modules and other transcriptional regulators.

processes including chromatin assembly[17], suppression of spurious transcriptional initiation[18] and the acetylation of histone H3 on lysine 56 (K56)[19].

Examination of interaction patterns suggests that the underlying functional connection between the members of this epistasis group is mediated through the cell-cycle-dependent acetylation of H3 K56 (refs 20–22). For example, deletion of *HST3*, which encodes the enzyme primarily responsible for H3 K56 deacetylation[21,22], shows positive genetic interactions with these genes (Fig. 4a). Consistent with this, deletion of *ASF1* (ref. 22), *RTT101*, *RTT109*, *MMS1* or *MMS22* suppresses the growth defect caused by hyperacetylation in *hst3*Δ *hst4*Δ strains (Fig. 4b). As seen for the *asf1*Δ strain, deletion of *RTT109* eliminates detectable K56 acetylation *in vivo* (Fig. 4c)[23]. Deletion or temperature-dependent inactivation of each known yeast histone acetyltransferase (HAT) protein has little effect on global levels of K56 acetylation *in vivo*[24], suggesting that Rtt109 itself might be the responsible HAT. Indeed, we found that Rtt109, both affinity-purified from yeast (Fig. 4d) and recombinantly produced from *Escherichia coli* (Fig. 4e), acetylates itself as well as recombinant histone H3. Western blot analysis using a residue-specific antibody confirmed that Rtt109 acts on K56 (Fig. 4d). Mutation of two universally conserved adjacent aspartic acid residues abolished K56 acetylation both *in vivo* and *in vitro* (Fig. 4f, g). Notably, Rtt109 is itself acetylated on K290 (data not shown), just two amino acids away from the conserved aspartic acids, suggesting that auto-acetylation may be a mechanism of regulation. Finally, addition of Asf1 markedly enhances the *in vitro* H3 K56 acetylation activity of Rtt109 (Supplementary Fig. 4), suggesting that the Asf1 requirement for the acetylation *in vivo* reflects Asf1's ability to bind to and present H3–H4 heterodimers[25,26]. These observations indicate that Rtt109 is the founding member of a new family of HATs that shares no detectable sequence similarities with previously known HATs.

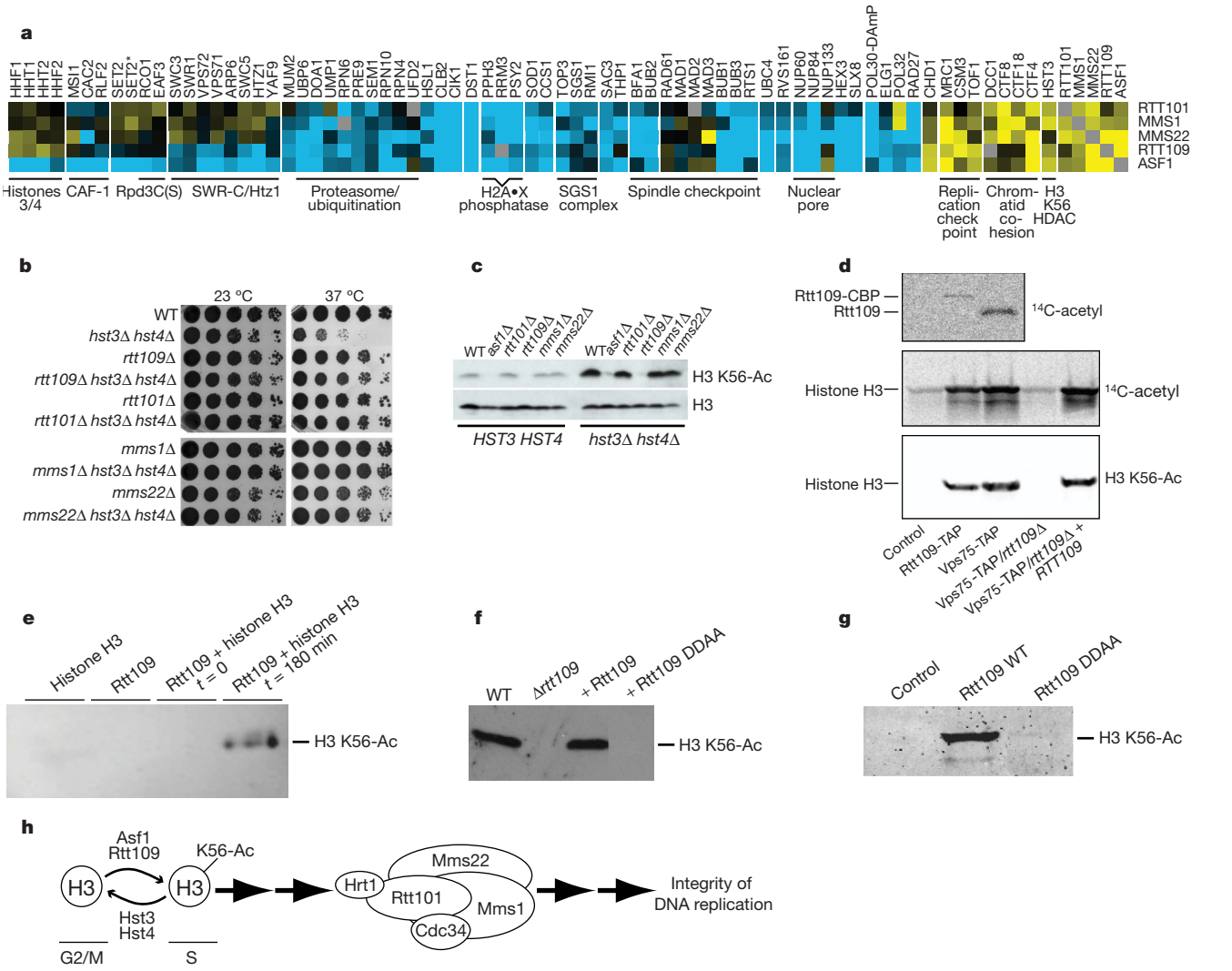


Figure 4 | Characterization of the histone H3 K56 acetylation pathway. **a**, A subset of the genetic interaction patterns for *RTT101*, *MMS22*, *MMS1*, *RTT109* and *ASF1*. Shading as in Fig. 1. **b**, The growth of strains was assessed by plating (sevenfold) serial dilutions and incubating at either 23 °C or 37 °C. **c**, Western blot analysis was used to assess the K56 acetylation levels in wild-type, *asf1*Δ, *rtt101*Δ, *rtt109*Δ, *mms1*Δ and *mms22*Δ strains in *HST3 HST4* (wild type) or *hst3*Δ *hst4*Δ backgrounds. **d**, *In vitro* acetylation activity of Rtt109 purified from either Rtt109-TAP or Vps75-TAP strains (Rtt109 and Vps75 are complexed (Supplementary Fig. 3)[8]) was monitored by incorporation of radiolabel from ^{14}C-acetyl-coenzyme A (CoA) into either the purified Rtt109 or recombinant *Xenopus laevis* histone H3 purified from bacteria. No protein was added in the control reaction. Histone acetylation was also monitored using antibody directed against K56 acetylated histone H3. CBP, calmodulin binding peptide. **e**, Recombinant histone H3, recombinant GST–Rtt109, or both, were incubated with acetyl-CoA. K56 acetylation was assessed by western blotting. **f**, Western blot analysis was used to assess K56 acetylation levels in a wild-type strain and in *rtt109*Δ strains episomally expressing either nothing (Δ*rtt109*), wild-type Rtt109 (+ Rtt109), or Rtt109 mutated at D287A and D288A (+ Rtt109 DDAA). **g**, *In vitro* HAT assays were carried out using the indicated Rtt109 protein purified from Vps75-TAP strains. Results were analysed by western blot analysis. Mass spectrometry analysis confirmed that mutated Rtt109 (DDAA) still physically associates with Vps75 (data not shown). **h**, Model for the histone H3 K56 acetylation signalling pathway.

Unlike Rtt109 and Asf1, loss of Rtt101, Mms22 or Mms1 suppresses the *HST3*/*HST4* double deletion without preventing H3 K56 hyperacetylation (Fig. 4b, c). This suggests that the Rtt101 cullin complex is the major downstream effector of this acetylation pathway, thereby providing insight into this modification's physiological role. Specifically, defects in the progression of replication forks through damaged DNA and natural pause sites[27] seen in a *rtt101*Δ strain (as demonstrated by prolonged activation of the checkpoint protein Rad53 after methyl methanesulphonate treatment) are also seen when H3 K56 acetylation is blocked by deletion of *RTT109* or *ASF1* (Supplementary Fig. 3). Furthermore, *RTT101*, *MMS22* and *MMS1* deletions confer no additional sensitivity to DNA-damaging agents when combined with deletion of *ASF1* (J. Erkmann and P. Kaufman, personal communication). Taken together, these data suggest a model (Fig. 4h) in which H3 K56 acetylation acts upstream of the Rtt101 ubiquitin ligase complex—which is presumably targeting one or more proteins for degradation by the proteasome—to promote replication fork progression through damaged DNA. On the basis of positive genetic interactions in the E-MAP (Fig. 4a; see also Supplementary Data), potential targets include components of the replication checkpoint complex (Mrc1, Csm3, Tof1) and proteins involved in chromatid cohesion (Dcc1, Ctf4, Ctf8, Ctf18). Given that H3 K56 acetylation accumulates to ~50% during S phase[22], K56 acetylation may generally serve to mark newly synthesized DNA and allow the activity of Rtt101 to be specifically targeted to such regions during DNA synthesis. More broadly, this example shows that analysis of E-MAPs can illuminate complex biological pathways with high precision. Success here and in earlier work using hypomorphic alleles[1,28] for genetic interaction studies points to the potential for extending this approach to metazoans using technologies such as RNA interference.

METHODS

Strains were constructed and E-MAP experiments were performed as described previously[29]. Genetic interaction scores were computed as described in ref. 6. Histone acetyltransferase assays were performed as described[30], using immunoprecipitates from whole-cell extracts prepared from yeast strains containing one gene with a tandem affinity purification (TAP) tag[8]. More details for experimental assays, as well as a description of the data analysis, are provided in Supplementary Methods.

Received 5 December; accepted 5 February 2007.
Published online 21 February 2007.

1. Schuldiner, M. *et al.* Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123,** 507–519 (2005).
2. Tong, A. H. *et al.* Global mapping of the yeast genetic interaction network. *Science* **303,** 808–813 (2004).
3. Pan, X. *et al.* A DNA integrity network in the yeast *Saccharomyces cerevisiae*. *Cell* **124,** 1069–1081 (2006).
4. Kelley, R. & Ideker, T. Systematic interpretation of genetic interactions using protein networks. *Nature Biotechnol.* **23,** 561–566 (2005).
5. Ye, P. *et al.* Gene function prediction from congruent synthetic lethal interactions in yeast. *Mol. Syst. Biol.* **1,** 2005.0026 (2005).
6. Collins, S. R., Schuldiner, M., Krogan, N. J. & Weissman, J. S. A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol.* **7,** R63 (2006).
7. Gavin, A. C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440,** 631–636 (2006).
8. Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440,** 637–643 (2006).
9. Collins, S. R. *et al.* Towards a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics* doi:10.1074/mcp.M600381-MCP200 (2 January 2007).
10. Hong, E. L. *et al.* *Saccharomyces* Genome Database ⟨http://www.yeastgenome.org/⟩ (7 March 2006).
11. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York, 2001).
12. Kornberg, R. D. Mediator and the mechanism of transcriptional activation. *Trends Biochem. Sci.* **30,** 235–239 (2005).
13. Bjorklund, S. & Gustafsson, C. M. The yeast Mediator complex and its regulation. *Trends Biochem. Sci.* **30,** 240–244 (2005).
14. van de Peppel, J. *et al.* Mediator expression profiling epistasis reveals a signal transduction pathway with antagonistic submodules and highly specific downstream targets. *Mol. Cell* **19,** 511–522 (2005).
15. Takagi, Y. *et al.* Head module control of mediator interactions. *Mol. Cell* **23,** 355–364 (2006).
16. Michel, J. J., McCarville, J. F. & Xiong, Y. A role for *Saccharomyces cerevisiae* Cul8 ubiquitin ligase in proper anaphase progression. *J. Biol. Chem.* **278,** 22828–22837 (2003).
17. Tyler, J. K. Chromatin assembly. Cooperation between histone chaperones and ATP-dependent nucleosome remodeling machines. *Eur. J. Biochem.* **269,** 2268–2274 (2002).
18. Schwabish, M. A. & Struhl, K. Asf1 mediates histone eviction and deposition during elongation by RNA polymerase II. *Mol. Cell* **22,** 415–422 (2006).
19. Recht, J. *et al.* Histone chaperone Asf1 is required for histone H3 lysine 56 acetylation, a modification associated with S phase in mitosis and meiosis. *Proc. Natl Acad. Sci. USA* **103,** 6988–6993 (2006).
20. Masumoto, H., Hawke, D., Kobayashi, R. & Verreault, A. A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. *Nature* **436,** 294–298 (2005).
21. Maas, N. L., Miller, K. M., DeFazio, L. G. & Toczyski, D. P. Cell cycle and checkpoint regulation of histone H3 K56 acetylation by Hst3 and Hst4. *Mol. Cell* **23,** 109–119 (2006).
22. Celic, I. *et al.* The sirtuins Hst3 and Hst4p preserve genome integrity by controlling histone H3 lysine 56 deacetylation. *Curr. Biol.* **16,** 1280–1289 (2006).
23. Schneider, J., Bajwa, P., Johnson, F. C., Bhaumik, S. R. & Shilatifard, A. Rtt109 is required for proper H3K56 acetylation: a chromatin mark associated with the elongating RNA polymerase II. *J. Biol. Chem.* **281,** 37270–37274 (2006).
24. Ozdemir, A. *et al.* Characterization of lysine 56 of histone H3 as an acetylation site in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **280,** 25949–25952 (2005).
25. Adkins, M. W., Carson, J. J., English, C. M., Ramey, C. J. & Tyler, J. K. The histone chaperone anti-silencing function 1 stimulates the acetylation of newly synthesized histone H3 in S-phase. *J. Biol. Chem.* **282,** 1334–1340 (2007).
26. English, C. M., Adkins, M. W., Carson, J. J., Churchill, M. E. & Tyler, J. K. Structural basis for the histone chaperone activity of Asf1. *Cell* **127,** 495–508 (2006).
27. Luke, B. *et al.* The cullin Rtt101p promotes replication fork progression through damaged DNA and natural pause sites. *Curr. Biol.* **16,** 786–792 (2006).
28. Davierwala, A. P. *et al.* The synthetic genetic interaction spectrum of essential genes. *Nature Genet.* **37,** 1147–1152 (2005).
29. Schuldiner, M., Collins, S. R., Weissman, J. S. & Krogan, N. J. Quantitative genetic analysis in *Saccharomyces cerevisiae* using epistatic miniarray profiles (E-MAPs) and its application to chromatin functions. *Methods* **40,** 344–352 (2006).
30. Mizzen, C. A., Brownell, J. E., Cook, R. G. & Allis, C. D. Histone acetyltransferases: preparation of substrates and assay procedures. *Methods Enzymol.* **304,** 675–696 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to K. Tipton and M. Bassik for critically reading the manuscript, S. Gasser, B. Frey and Vincent Cheung for discussion, and G. Narlikar for reagents. We thank N. Datta, T. Punna, N. Thompson, M. Ballantine, N. Gabovic, A. Wind, K. Chin, Y. Xue, A. Chan, Y. Xue, T. Chan, M. Xan, M. Lim, H. Dalgleish, K. Vachon, L. Le, C. Sun, Z. Hassam, J. Rilestone and K. Takhar for technical assistance. We also thank S. Jackson, Z. Zhang, Vanessa Cheung, F. Winston, J. Erkmann and P. Kaufman for communicating results before publication. This research was supported by grants from Genome Canada and the Ontario Genomics Institute (J.F.G., A.E., C.B. and B.A.), the NIH (D.P.T.), the Howard Hughes Medical Institute (J.S.W.) and the Canadian Institute of Health Research (N.J.K., C.J.I. and G.W.B.). S.R.C. was funded by a fellowship from the Burroughs Wellcome Fund. N.J.K. is a Sandler Family Fellow.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.S.W. (weissman@cmp.ucsf.edu) or J.F.G. (jack.greenblatt@utoronto.ca).

LETTERS

Conserved factors regulate signalling in *Arabidopsis thaliana* shoot and root stem cell organizers

Ananda K. Sarkar[1]*†, Marijn Luijten[2]*, Shunsuke Miyashima[3], Michael Lenhard[1]†, Takashi Hashimoto[3], Keiji Nakajima[3], Ben Scheres[2], Renze Heidstra[2] & Thomas Laux[1]

Throughout the lifespan of a plant, which in some cases can last more than one thousand years, the stem cell niches in the root and shoot apical meristems provide cells for the formation of complete root and shoot systems, respectively. Both niches are superficially different and it has remained unclear whether common regulatory mechanisms exist. Here we address whether root and shoot meristems use related factors for stem cell maintenance. In the root niche the quiescent centre cells, surrounded by the stem cells, express the homeobox gene *WOX5* (*WUSCHEL-RELATED HOMEOBOX 5*)[1], a homologue of the *WUSCHEL* (*WUS*) gene that non-cell-autonomously maintains stem cells in the shoot meristem[2]. Loss of *WOX5* function in the root meristem stem cell niche causes terminal differentiation in distal stem cells and, redundantly with other regulators, also provokes differentiation of the proximal meristem. Conversely, gain of *WOX5* function blocks differentiation of distal stem cell descendents that normally differentiate. Importantly, both *WOX5* and *WUS* maintain stem cells in either a root or shoot context. Together, our data indicate that stem cell maintenance signalling in both meristems employs related regulators.

Higher organisms evolved the ability to keep founder cells undifferentiated and pluripotent by signals provided in specialized stem cell niches[3]. In the model plant *Arabidopsis thaliana*, root and shoot meristem stem cell niches are organized differently compared with each other[4]. In the root meristem, the stem cells surround a small group of rarely dividing cells, termed the quiescent centre (Fig. 1a), and give rise to distal (columella), lateral (lateral root cap and epidermis) and proximal (cortex, endodermis and stele) cell types. Ablation studies show that short range signals from the quiescent centre keep only the directly abutting stem cells undifferentiated[5], similar to most animal stem cell niches studied so far[3]. In contrast, in the shoot meristem, a zone of three stem cell layers is maintained by an underlying organizing centre[2] (Supplementary Fig. 1a), which regulates the stem cell pool as a whole. In addition, the regulatory genes described so far for root and shoot stem cell niches are different[6].

In the shoot meristem, a feedback mechanism between the organizing centre and stem cells dynamically regulates maintenance of the stem cell pool: *WUS* activity in the organizing centre keeps stem cells undifferentiated and induces expression of the signal peptide CLAVATA3 (CLV3), which in turn restricts the size of the *WUS* expression domain (Supplementary Fig. 1a)[7,8]. We asked whether a related stem cell maintenance mechanism might operate in the root by analysing the *WOX* family for functions in the root meristem[1]. Expression of *WOX5* messenger RNA[1] and a *WOX5-GUS* reporter gene initiates in the embryonic cell lineage that gives rise to the quiescent centre (Supplementary Fig. 1b), and persists in the quiescent centre (Fig. 1b; and Supplementary Fig. 1c) during post-embryonic root growth. This expression pattern is strikingly similar to that of *WUS* in the shoot meristem[2] (Supplementary Fig. 2), raising the question of whether both genes might have similar functions in the respective stem cell niches.

In seedlings of the putative null allele *wox5-1* (Supplementary Fig. 1d, e), the cells at the quiescent centre position have abnormal shape and are enlarged in comparison to wild type (Fig. 1c, d, arrows, and Supplementary Table 1). Notably, adjacent columella stem cells are even larger, suggesting that they have undergone differentiation (Fig. 1d, bottom arrow). No other abnormality was observed in *wox5-1* mutants. Complementation of the mutant defects by a *WOX5* complementary DNA construct confirmed that the *wox5-1* mutation causes the observed phenotype (Supplementary Fig. 3a).

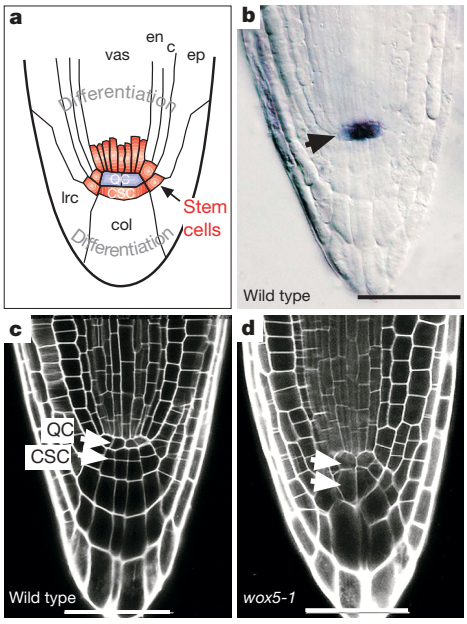


Figure 1 | *WOX5* expression in the quiescent centre is required for root meristem development. a, Schematic of *A. thaliana* root meristem with columella (col), columella stem cells (CSC), quiescent centre (QC), lateral root cap (lrc), epidermis (ep), cortex (c), endodermis (en) and vascular bundle (vas), redrawn with permission from ref. 4. **b**, *WOX5* mRNA expression in the quiescent centre. **c–d**, *wox5-1* roots display enlarged cells at the quiescent centre and columella stem cell (CSC) positions. Scale bars, 50 μm.

[1]Institute of Biology III, University of Freiburg, Schänzlestraße 1, 79104 Freiburg, Germany. [2]Department of Molecular Cell Biology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands. [3]Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan. †Present addresses: Cold Spring Harbor Laboratory, One Bungtown Road, Cold Spring Harbor, New York 11724, USA (A.K.S.); John Innes Centre, Colney Lane, Norwich, N4 7UH, UK (M.K.).
*These authors contributed equally to this work.

Of several quiescent-centre-specific reporter genes[9], only *QC184* is not expressed in *wox5-1* and expression is lacking as early as the heart embryo stage (Fig. 2a, b, d, e; and Supplementary Table 2), indicating that *WOX5* functions in root development from early embryogenesis. In contrast, *QC25*, *QC46* and *WOX5-GUS* are expressed in slightly expanded domains compared to wild type (Fig. 2c, f; Supplementary Fig. 3c–h; and Supplementary Table 2). These data indicate that *WOX5* is required for some aspects of quiescent-centre-specific gene expression, but seems not to be a major factor in quiescent centre specification.

We asked whether in *wox5-1* roots the distal stem cells might have undergone premature differentiation, which in the columella can be visualized by the accumulation of starch-granule-containing organelles involved in gravitropic sensing. Indeed, in *wox5-1* roots, the cells at the stem cell position, but not the quiescent centre, accumulate starch granules (Fig. 2e, f; and Supplementary Table 2). In contrast, the proximal meristem appears normal. These findings indicate that *WOX5* gene expression in the quiescent centre is required to non-cell-autonomously maintain the distal stem cells as undifferentiated, similar to the role of *WUS* in the shoot[2].

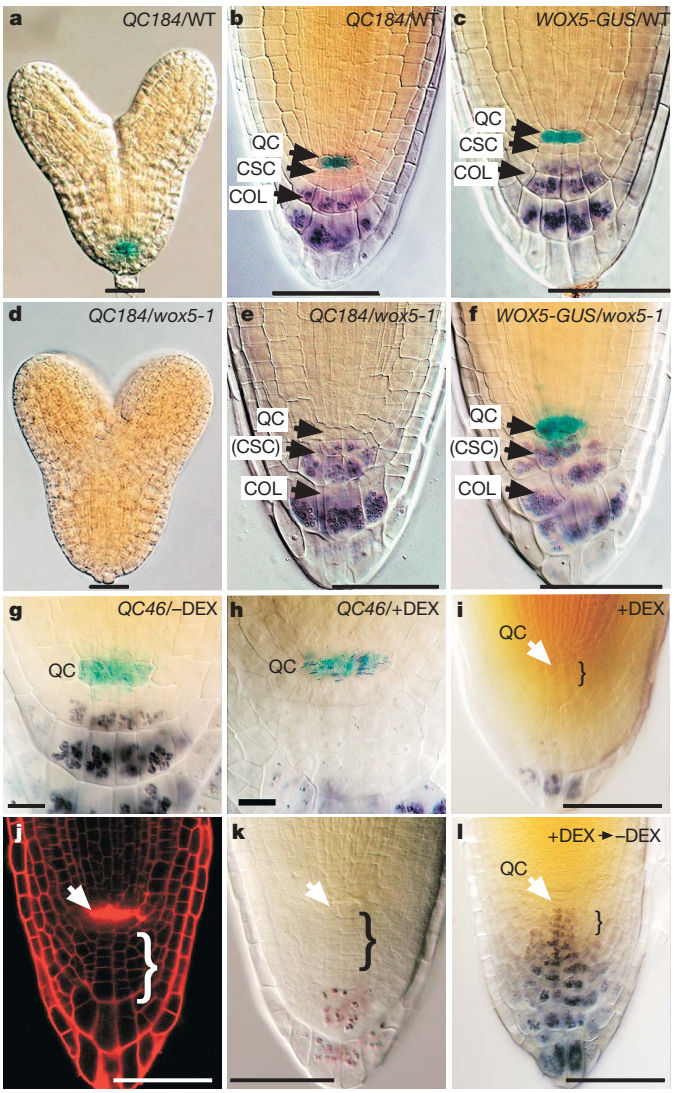


Figure 2 | *WOX5* represses differentiation in the columella. **a–f**, *wox5-1* heart-torpedo stage embryos (**d**) and root tips (**e**) lack *QC184* expression, express *WOX5-GUS* (**f**) in an enlarged domain and accumulate starch at the CSC position (**e**, **f**). **g**, **h**, *35S-WOX5* expression gives small cells in the columella, which lack *QC46* expression and starch grains (**h**). **j–l**, Quiescent centre ablation (arrow) does not induce differentiation (**j**, **k**), whereas transfer to non-inducing medium does (**l**; control, **i**). Nomarski (**a–i**, **k**, **l**) and confocal (**j**) median longitudinal images are shown. GUS staining, blue; starch, violet; waved brackets, extra small cells; COL, columella; CSC, columella stem cells; QC, quiescent centre; DEX, dexamethasone. Scale bars, 20 μm (**a**, **d**), 50 μm (**b**, **c**, **e**, **f**, **i–l**) and 10 μm (**g**, **h**).

We then investigated whether *WOX5* activity is also sufficient to repress columella cell differentiation, by ubiquitously expressing a dexamethasone (DEX)-inducible *WOX5* gene. DEX induction results in small cells that lack starch granules in the place of differentiated columella and lateral root cap cells (Fig. 2h, i; Supplementary Fig. 4a; and Supplementary Table 1), indicating that they have not undergone normal differentiation. Unlike quiescent centre cells, the small cells divide, as shown by the continuously increasing cell number and expression of the cell cycle marker *CYCB1-GUS* (Supplementary Fig. 4b) and quiescent centre markers are not induced (Fig. 2h). Consistent with a loss of differentiated columella cells, *35S[GVG]-WOX5* (*35S-WOX5*; Supplementary Methods) roots display reduced gravitropism (Supplementary Fig. 4c). After withdrawal of DEX, however, the small cells readily accumulate starch grains typical for differentiated columella cells (Fig. 2l). Normally, laser ablation of the quiescent centre causes differentiation of stem cells[9,10], but this did not have any effect in *35S-WOX5* (Fig. 2j, k). Thus, ectopic *WOX5* expression is sufficient to block differentiation of columella stem cell daughters without the requirement of any further quiescent centre signalling.

Previous studies highlighted auxin-dependent transcription of the *PLETHORA* (*PLT*) genes and *SHORTROOT/SCARECROW* (*SHR/SCR*) functions as essential components in root meristem maintenance and we asked how *WOX5* interacts with these pathways[9,11–13]. In mutants for the auxin response regulators *BODENLOS* (*BDL*) or *MONOPTEROS* (*MP;* Fig. 3d–f; and Supplementary Table 3), *WOX5* expression is rarely detected, consistent with the notion that *BDL*/*MP*-mediated auxin signalling is required for the embryonic initiation of the quiescent centre[13,14]. Auxin accumulation in the quiescent centre region provides patterning information to the root meristem mediated through *PLT* genes[9,12,15]. In *wox5-1* mutants, auxin accumulation seems normal based on the expression of the auxin response reporter gene *DR5-GUS* (Supplementary Fig. 5a, b). In *plt1 plt2* double mutants, *WOX5* expression is occasionally slightly expanded (Fig. 3g–i; and Supplementary Table 3), whereas *PLT1* expression is normal in *wox5-1* (Supplementary Fig. 5c, d and Supplementary Table 2).

The SHR protein moves from provascular cells into the quiescent centre and activates transcription of the *SCR* gene[16], which is required to specify quiescent centre identity and to maintain stem cells[11]. *WOX5* expression is reduced or undetectable in *shr* and *scr* mutants (Fig. 3j–l; Supplementary Fig. 5g–i; and Supplementary Table 3), whereas an *SCR* reporter gene is correctly expressed in *wox5-1* roots (Supplementary Fig. 5e, f; and Supplementary Table 2). Collectively, our data indicate that *WOX5* expression depends on the induction in the root pole by *MP*-mediated auxin signalling and on *SHR/SCR* activity, whereas *PLT1* and *PLT2* have only a minor role in confining *WOX5* expression to the quiescent centre.

The requirement of *WOX5* for only columella stem cell maintenance suggests unidirectionality of this signalling pathway in stem cell control, similar to *WUS*-mediated stem cell control in the shoot stem cell niche. We investigated whether effects of the *wox5-1* mutation on the proximal meristem might be masked by redundant functions. Indeed, in *scr-4*, *shr-1* single mutants and *plt1 plt2* double mutants, the *wox5-1* mutation promotes differentiation in the proximal region of the root meristem in addition to its effects on distal stem cells (Fig. 3m; and Supplementary Fig. 5j, k). This suggests that *WOX5* gene expression in the quiescent centre redundantly contributes to proximal stem cell activity or alternatively that *WOX5* has a stem-cell-independent function in the proximal meristem, as has been reported for its upstream regulator SCR[11] (Supplementary Fig. 7).

The striking similarities of *WOX5* and *WUS* expression patterns and functions in root and shoot stem cell niche organizers raised the

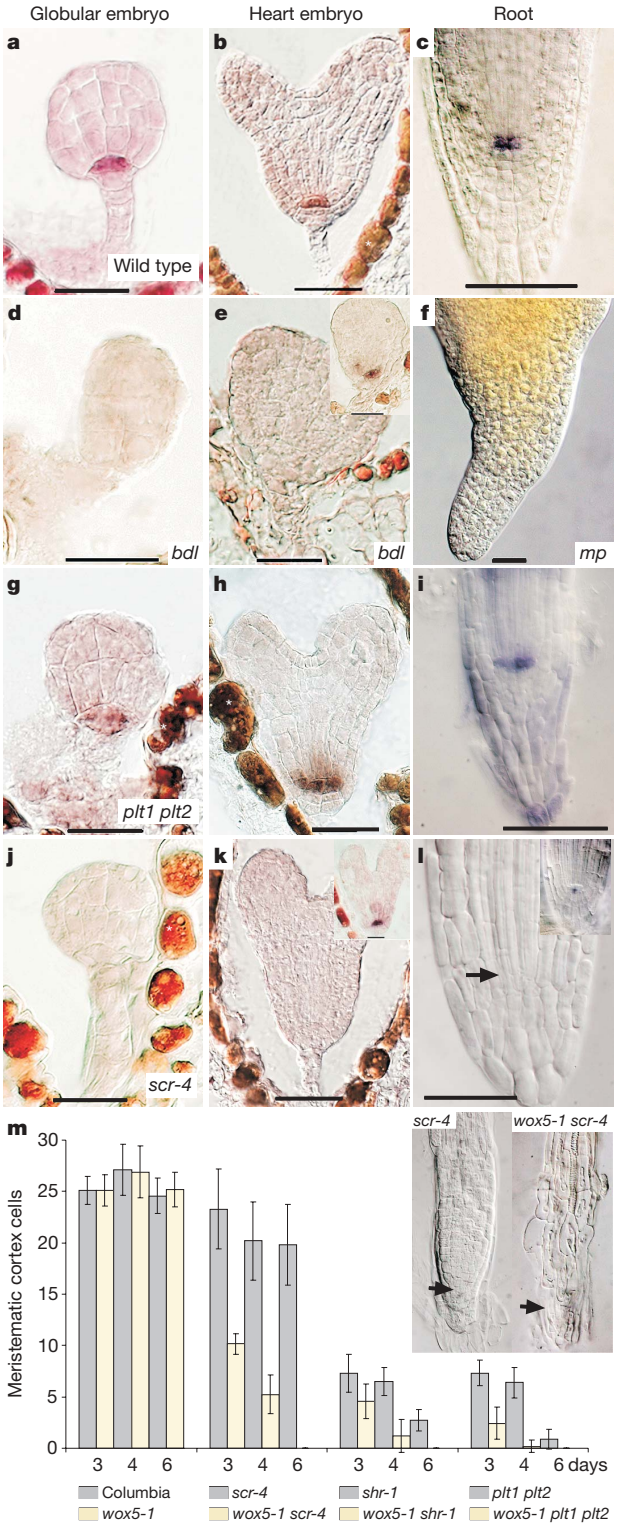


Figure 3 | *WOX5* interaction with *SHR/SCR* and auxin pathways. **a**–**l**, Hybridization with *WOX5* antisense probe. *WOX5* expression was detected in wild type (**a**–**c**), absent in *bdl* and *mp* (**d**–**f**, inset shows sporadic exceptions), deregulated in *plt1 plt2* (**g**–**i**), and undetectable in *scr-4* (**j**–**l**, insets show sporadic exceptions). **m**, *wox5-1* enhances differentiation in the proximal meristem in *scr-4* (inset), *shr-1* and *plt1 plt2* mutants. Error bars, s.d. Longitudinal median histological (**a**, **b**, **d**, **e**, **g**, **h**, **j**, **k**) and optical (**c**, **f**, **i**, **l**, **m**) sections are shown. Four-day-old roots (**c**, **i**, **l**) or middle stage embryos (**f**) were used. Hybridization signal is red–blue; asterisk, reddish colour is staining independent of hybridization reaction. Arrows, quiescent centre position. Scale bars, 25 μm (**a**, **b**, **d**–**f**, **g**, **h**, **j**, **k**) and 50 μm (**c**, **i**, **l**).

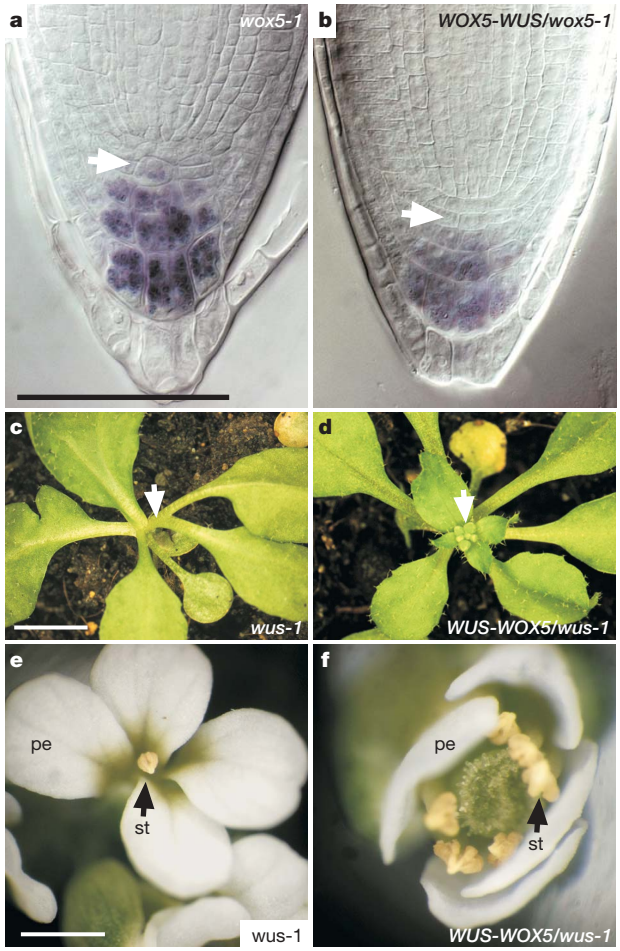


Figure 4 | *WOX5* and *WUS* are interchangeable in stem cell control. **a**, **b**, *WOX5-WUS* expression restores quiescent centre and columella stem cells (white arrows) in *wox5-1*. Violet, starch grains. **c**–**f**, Rescue of the shoot meristem stem cell niche. *wus-1* shoot (**c**) and floral meristem (**e**) fail to maintain stem cells and terminate prematurely. *WUS-WOX5* expression restores an indeterminate inflorescence meristem (**d**) and complete flowers (**f**). Twenty-six-day-old plants (**c**, **d**) and flowers (**e**, **f**) are shown. White arrows indicate shoot apex (**c**, **d**). pe, petal; st, stamen. Scale bars, 1 cm (**c**, **d**), 1 mm (**e**, **f**) and 50 μm (**a**, **b**).

question of whether both genes are functionally equivalent. Indeed, expression of the *WUS* cDNA from the *WOX5* promoter completely restores quiescent centre and stem cells in the *wox5-1* root meristem (Fig. 4a, b; and Supplementary Fig. 6a, b). Interestingly, ectopic shoot tissue was never observed when *WUS* was expressed in the quiescent centre. This contrasts to previous studies expressing *WUS* in a broader range of cell types[17] and favours the interpretation that organizer signalling only maintains stem cells as undifferentiated, whereas the fate of differentiating daughters is determined by the tissue context. Conversely, expression of a *WUS-WOX5* transgene rescues premature termination of inflorescence meristems and occasionally of floral meristems in the putative null allele *wus-1* (Fig. 4c–f; and Supplementary Fig. 6d–g). The rescued *wus-1* mutants resemble the weak *wus-6/jam* allele[18], suggesting that the *WUS-WOX5* transgene complements *WUS* function but provides a reduced level of organizing centre signalling. Defective ovule development in *wus* mutants is not rescued by *WUS-WOX5* (Supplementary Fig. 6c), indicating that both regulators are interchangeable in only stem cell maintenance. Significantly, the early embryo genes *WOX8* and *WOX9* are not able to rescue *wus-1* defects (H. Breuninger and T.L., unpublished observations), showing that the ability to regulate stem cell maintenance is not a general property of WOX proteins.

The extent of mechanistic similarity and evolutionary relationship between root and shoot stem cell niches has remained unclear thus far. Here we show that the organizers of both niches employ related regulators to provide stem cell maintaining signals. However, there also seem to be differences between both niches. For example, *CLV3* homologues promote meristem differentiation but do not influence stem cell maintenance in the root on overexpression and it is thus unknown whether roots, similar to shoots, employ a *WUS–CLV3* related feedback mechanism[19,20].

Our results suggest that *WOX5* in the root stem cell niche has a more direct function in stem cell signalling, rather than in specifying quiescent centre identity (Supplementary Fig. 7). The nature of the signals that are induced by *WUS*/*WOX5* expression in the stem cell organizers is presently unresolved. WOX5 protein itself or a downstream factor might move to stem cells as the long-postulated short-range factor[5]. To date, WOX5 protein has not been robustly detected, presumably owing to very low expression levels. However, on the basis of localization of a functional WUS–GUS fusion protein, movement of WUS protein out of the organizing centre seems not to be necessary for its function in shoot meristem stem cell maintenance, indicating that factors downstream of *WUS* act as signals (Supplementary Fig. 6h, i).

Palaeobotanical evidence for early root structures is controversial, but the current view is that roots have evolved independently at least twice in Lycophytina and Euphyllophytina[21,22]. It remains to be shown whether *WUS* and *WOX5* shared an ancestral function in stem cell control since root and shoot separation, or whether they have been recruited for this function after the diversification of stem cell niches.

METHODS

Plant work. *wox5-1* (SALK038262) and *wox5-3* (SALK147644) are transfer (T)-DNA insertion alleles in the Columbia (Col) accession and were obtained from the *Arabidopsis* Biological Resource Center (ABRC, USA) and the Nottingham *Arabidopsis* Stock Center (NASC, UK). *7×DR5-GUS*[23] was kindly provided by J. Marfett, and mp^{U51} (ref. 14) by G. Jürgens (Tübingen). Root length and meristem size were measured as described[10]. The number of meristematic cells was obtained by counting cortex cells showing no signs of rapid elongation. Quiescent centre laser ablations were performed in 4-day-old roots using a Leica SP2 inverted confocal laser scanning microscope[5]. Roots were stained for amyloplasts using Lugol (Sigma) 24-h after ablation[24]. For confocal microscopy, root tips were mounted in 10 μg ml^{-1} propidium iodide solution. Microscope settings were as described[5].

Expression analysis. Starch granules[24] and β-glucuronidase activity[7] were visualized as described. *35S-WOX5* seedlings were cultured on standard 0.5× MS medium supplemented with or without 1 μM dexamethasone. *In situ* hybridization of sections, whole-mount embryos, and 4-day-old seedlings were performed as described[1,24]. The *PLT1* riboprobe has been described previously[12]. The *WOX5* riboprobe was prepared from a cDNA without the homeobox to avoid cross-hybridization with related mRNAs.

Further experimental details are provided as Supplementary Information.

Received 9 January; accepted 23 February 2007.

1. Haecker, A. *et al.* Expression dynamics of *WOX* genes mark cell fate decisions during early embryonic patterning in *Arabidopsis thaliana*. *Development* **131,** 657–668 (2004).
2. Mayer, K. F. X. *et al.* Role of *WUSCHEL* in regulating stem cell fate in the *Arabidopsis* shoot meristem. *Cell* **95,** 805–815 (1998).
3. Spradling, A., Drummond-Barbosa, D. & Kai, T. Stem cells find their niche. *Nature* **414,** 98–104 (2001).
4. Laux, T. The stem cell concept in plants: a matter of debate. *Cell* **113,** 281–283 (2003).
5. van den Berg, C., Willemsen, V., Hendriks, G., Weisbeek, P. & Scheres, B. Short-range control of cell differentiation in the *Arabidopsis* root meristem. *Nature* **390,** 287–289 (1997).
6. Vernoux, T. & Benfey, P. N. Signals that regulate stem cell activity during plant development. *Curr. Opin. Genet. Dev.* **15,** 388–394 (2005).
7. Schoof, H. *et al.* The stem cell population of *Arabidopsis* shoot meristems is maintained by a regulatory loop between the *CLAVATA* and *WUSCHEL* genes. *Cell* **100,** 635–644 (2000).
8. Brand, U., Fletcher, J. C., Hobe, M., Meyerowitz, E. M. & Simon, R. Dependence of stem cell fate in *Arabidopsis* on a feedback loop regulated by *CLV3* activity. *Science* **289,** 617–619 (2000).
9. Sabatini, S. *et al.* An auxin-dependent distal organizer of pattern and polarity in the *Arabidopsis* root. *Cell* **99,** 463–472 (1999).
10. Wildwater, M. *et al.* The *RETINOBLASTOMA-RELATED* gene regulates stem cell maintenance in *Arabidopsis* roots. *Cell* **123,** 1337–1349 (2005).
11. Sabatini, S., Heidstra, R., Wildwater, M. & Scheres, B. SCARECROW is involved in positioning the stem cell niche in the *Arabidopsis* root meristem. *Genes Dev.* **17,** 354–358 (2003).
12. Aida, M. *et al.* The *PLETHORA* genes mediate patterning of the *Arabidopsis* root stem cell niche. *Cell* **119,** 109–120 (2004).
13. Weijers, D. *et al.* Auxin triggers transient local signaling for cell specification in *Arabidopsis* embryogenesis. *Dev. Cell* **10,** 265–270 (2006).
14. Berleth, T. & Jürgens, G. The role of the *Monopteros* gene in organising the basal body region of the *Arabidopsis* embryo. *Development* **118,** 575–587 (1993).
15. Blilou, I. *et al.* The PIN auxin efflux facilitator network controls growth and patterning in *Arabidopsis* roots. *Nature* **433,** 39–44 (2005).
16. Nakajima, K., Sena, G., Nawy, T. & Benfey, P. Intercellular movement of the putative transcription factor SHR in root patterning. *Nature* **413,** 307–311 (2001).
17. Gallois, J. L., Nora, F. R., Mizukami, Y. & Sablowski, R. WUSCHEL induces shoot stem cell activity and developmental plasticity in the root meristem. *Genes Dev.* **18,** 375–380 (2004).
18. Hamada, S. *et al.* Mutations in the *WUSCHEL* gene of *Arabidopsis thaliana* result in the development of shoots without juvenile leaves. *Plant J.* **24,** 91–101 (2000).
19. Casamitjana-Martinez, E. *et al.* Root-specific *CLE19* overexpression and the *sol1/2* suppressors implicate a CLV-like pathway in the control of *Arabidopsis* root meristem maintenance. *Curr. Biol.* **13,** 1435–1441 (2003).
20. Hobe, M., Muller, R., Grunewald, M., Brand, U. & Simon, R. Loss of CLE40, a protein functionally equivalent to the stem cell restricting signal CLV3, enhances root waving in *Arabidopsis*. *Dev. Genes Evol.* **213,** 371–381 (2003).
21. Kenrick, P. & Crane, P. R. The origin and early evolution of plants on land. *Nature* **389,** 33–39 (1997).
22. Kenrick, P. & Crane, P. *The Origin and Early Diversification of Land Plants: a Cladistic study* (Smithsonian Institution Press, Washington, 1997).
23. Ulmasov, T., Hagen, G. & Guilfoyle, T. J. ARF1, a transcription factor that binds to auxin response elements. *Science* **276,** 1865–1868 (1997).
24. Willemsen, V., Wolkenfelt, H., de Vrieze, G., Weisbeek, P. & Scheres, B. The *HOBBIT* gene is required for formation of the root meristem in the *Arabidopsis* embryo. *Development* **125,** 521–531 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to E. van der Graaff for the data shown in Supplementary Fig. 4c, and E. Tucker, T. Demlow and B. Geiges for experimental help. We thank G. Jürgens for *mp* seeds, E. Kiegle and J. Haseloff for mGFP5ER and UAS vectors, M. Terlou for software development for root and meristem measurements, P. Graf for photography and E. Tucker and M. Tucker for comments on the manuscript. This work was sponsored by grants from the European Union (REGIA) and the Deutsche Forschungsgemeinschaft (to T.L.), a Netherlands Genomics Initiative grant (to R.H.), and grants from the Japan Society for the Promotion of Science and the Novartis Foundation (to K.N.).

Author Information Sequence of *WOX5* mRNA is available in GenBank under accession numbers NM_111961 and AY251398. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.L. (laux@biologie.uni-freiburg.de).

LETTERS

Synthetic lethal screen identification of chemosensitizer loci in cancer cells

Angelique W. Whitehurst[1], Brian O. Bodemann[1], Jessica Cardenas[1], Deborah Ferguson[2], Luc Girard[3], Michael Peyton[3], John D. Minna[3,4], Carolyn Michnoff[5], Weihua Hao[5], Michael G. Roth[5], Xian-Jin Xie[4,6] & Michael A. White[1,4]

Abundant evidence suggests that a unifying principle governing the molecular pathology of cancer is the co-dependent aberrant regulation of core machinery driving proliferation and suppressing apoptosis[1]. Anomalous proteins engaged in support of this tumorigenic regulatory environment most probably represent optimal intervention targets in a heterogeneous population of cancer cells. The advent of RNA-mediated interference (RNAi)-based functional genomics provides the opportunity to derive unbiased comprehensive collections of validated gene targets supporting critical biological systems outside the framework of preconceived notions of mechanistic relationships. We have combined a high-throughput cell-based one-well/one-gene screening platform with a genome-wide synthetic library of chemically synthesized small interfering RNAs for systematic interrogation of the molecular underpinnings of cancer cell chemoresponsiveness. NCI-H1155, a human non-small-cell lung cancer line, was employed in a paclitaxel-dependent synthetic lethal screen designed to identify gene targets that specifically reduce cell viability in the presence of otherwise sublethal concentrations of paclitaxel. Using a stringent objective statistical algorithm to reduce false discovery rates below 5%, we isolated a panel of 87 genes that represent major focal points of the autonomous response of cancer cells to the abrogation of microtubule dynamics. Here we show that several of these targets sensitize lung cancer cells to paclitaxel concentrations 1,000-fold lower than otherwise required for a significant response, and we identify mechanistic relationships between cancer-associated aberrant gene expression programmes and the basic cellular machinery required for robust mitotic progression.

Paclitaxel and related taxanes are routinely used in the treatment of non-small-cell lung cancer (NSCLC) and other epithelial malignancies. Although objective responses and survival benefits are seen, complete responses are uncommon. Half-maximal inhibitory concentrations (IC_{50} values) of Paclitaxel in a panel of 29 human primary lung-tumour-derived cell lines spanned a wide range of concentrations, from 1 nM to more than 1 mM (M.P., L.G. and J.D.M., unpublished observations). From this panel, we selected the NSCLC line NCI-H1155 for a genome-wide paclitaxel synthetic-lethal screen, given the IC_{50} (about 50 nM) for this line, which is tenfold that observed for many other lines with similar proliferation rates. A high-efficiency, high-throughput short interfering RNA (siRNA) reverse transfection protocol was designed on the basis of our observations that transient trypsin-mediated suspension of adherent cultures markedly enhances the cellular uptake of liposome/nucleic acid particles[2,3] (Supplementary Fig. 1a).

We employed a library of 84,508 siRNAs corresponding to four unique siRNA duplexes, targeting each of 21,127 unique human genes arrayed in a one-gene–one-well format on 96-well microtitre plates (Supplementary Table 1). Transfections were performed in sextuplicate for triplicate analysis in the presence and the absence of paclitaxel (Supplementary Fig. 1b). A 48-h exposure to 10 nM paclitaxel was used as an otherwise innocuous dose that was in range of a significant response at a tenfold higher drug concentration (Supplementary Fig. 1c). Cell viability was measured using ATP concentration, and raw values were normalized to internal reference control samples on each plate to permit plate-to-plate comparisons[4] (Supplementary Table 1 and Supplementary Fig. 1f). Each siRNA pool was assigned a viability ratio calculated as mean viability in paclitaxel divided by mean viability in the absence of drug ($mean_{paclitaxel}/mean_{carrier}$) (Supplementary Fig. 1d).

An objective protocol for the selection of significant 'hits' was derived to combine reproducibility of testing with magnitude of response (Supplementary Fig. 1e, and Methods described therein). First, we set a 5% false discovery rate (FDR) by using two-sample *t*-tests from the triplicate analysis together with *P*-value corrections for the multiplicity of testing[5–8] (Supplementary Table 2). Second, we selected all samples that both satisfied a 5% FDR and were present in the 2.5-centile rank of the viability ratios (Supplementary Table 3 and Supplementary Fig. 1e). A set of 87 candidate paclitaxel-sensitizer loci, defined as 'high-confidence hits', was identified that satisfied these criteria (Table 1). Retests of a subset of these candidates with independently synthesized siRNA pools reproduced a significant response to 10 nM paclitaxel (Fig. 1a).

A 5% FDR is a highly stringent cut-off that may produce many false negatives. Nonetheless, this cut off returned many candidates with overlapping functional relationships, including macromolecular complexes, receptor–ligand pairs, and products of related aberrant gene-expression programmes (Table 1, Fig. 1b–e, and Supplementary Table 4). Most striking was the presence of a large group of core components of the proteasome (Fig. 1b), consistent with numerous empirical observations of enhanced sensitivity to paclitaxel in cancer cells after proteasome inhibition. Multiple targets encoding proteins involved in the dynamics and function of microtubules were also isolated[9] (Table 1). Relaxing the FDR to 10% returned most of the known main components of the γ-tubulin ring complex (γ-TuRC), a central element of the microtubule organizing centres that nucleate the formation of the mitotic spindle[10] (Fig. 1c, and Supplementary Table 4). Isolation of these components is evocative of a successful primary screen, because the mitotic spindle apparatus is exquisitely sensitive to the inhibition of microtubule dynamics by paclitaxel and is probably the biologically relevant drug target in cancer cells[11]. The probabilities of this extent of enrichment of proteasome subunits and of γ-TuRC subunits by random chance are 1 in 10^{10} ($P \leq 0.0000000001$)

[1]Department of Cell Biology, [2]Reata Pharmaceuticals, [3]Hamon Center for Therapeutic Oncology Research, [4]Simmons Cancer Center, [5]Department of Biochemistry, and [6]Center for Biostatistics and Clinical Science, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA.

Table 1 | High-confidence hit list

Symbol	Comments; motifs	Symbol	Comments; motifs
Proteasome		Transcription	
PSMA6	Proteasome subunit	RP9	ZnF_C2HC
PSMA7	Proteasome subunit	ZFPM1	ZnF_C2H2(x9)
PSMA8 (MGC26605)	Proteasome subunit	ZNF503	ZnF_C2H2
PSMB1	Proteasome subunit	ZNF585A	KRAB; ZnF_C2H2(x21)
PSMC3	Proteasome subunit	C11ORF30	ENT
PSMD1	Proteasome subunit	TRIM15	RING, BBOX, PRY, SPRY
PSMD3	Proteasome subunit		
		Translation	
Microtubule-related		RARSL	Arginyl-tRNA synthetase-like; Arg_S Core, tRNA-synt_1d_C
TUBGCP2	γ-TURC subunit; Spc97_Spc98	LOC390876	Similar to 60S ribosomal protein L35; coiled-coil
TUBA8	α-Tubulin	LOC388568	Similar to ribosomal protein S15 isoform
DNHD1 (FLJ32752)	Dynein heavy-chain subunit	SYMPK	
DNAH10 (FLJ43808)	Dynein heavy-chain subunit	SYNCRIP	RRM
TBL1Y	Transducin (β)-like 1Y-linked; LisH, WD40	BCDIN3 (FLJ20257)	Bin3, PrmA
MPP7	MAGUK family; L27, PDZ_signalling, SH3, GMPK	LOC144233	Bin3
Post-translational modification		Channel	
FBXO18	FBOX, UvrD-helicase	ATP6V0D2	Lysosomal H^+ transporter; vATP-synt_AC39
RAI17	Similar to PIAS; zf-MIZ	SLC34A3	Solute carrier; Na_Pi_cotrans
RNF151	RING		
LOC389822 (DKFZp434E1818)	Transmembrane, RING	Membrane protein	
LOC401506 (new LOC648245)	RING	BEAN	Transmembrane, basic domain
HS6ST2	Heparan sulphotransferase	LRRTM1	LLRNT, LRR (×9), transmembrane
GAL3ST4	Galactose sulphotransferase	MGC31967	Transmembrane, C-C
MGC4655	Galactosyl_T	TMC05 (MGC35118)	Transmembrane, C-C
Cell adhesion/ECM receptor		Other	
PAPLN	Proteoglycan-like sulphated glycoprotein; TSP1.KU, IGcam	PDDC1 (FLJ34283)	GATase1_Hsp31_like
KIAA1920	Similar to chondroitin sulphate proteoglycan 4	C14ORF148	Predicted NADP oxidoreductase; P5CR
LRFN5	LRR, COG4886, IGcam, transmembrane	CWF19L2	Similar to CWF19; CwfJ_C_1, CwfJ_C_2
MGC33424	IG_FLMN, CAP10, KDEL	PRICKLE1	Planar cell polarity, nuclear receptor for REST; PET, LIM (×3)
C1QTNF3	C1q	CA10	Carbonic anhydrase; alpha_CARP_X_XI_like
ITIH5	VIT, vWA_interalpha_trypsin_inhibitor	FAM14B	Similar to ISG12, aIFN-inducible; Ifi-6-16
IGSF21 (MGC15730)	IGcam	TIP39	PTHR2 ligand
		HSN2	Hereditary sensory neuropathy locus
Gametogenesis-associated/cancer testis antigens			
ACRBP	sp32, Kazal	Unknown	
FMR1NB	Transmembrane (×2), basic domain	NOD9	NACHT, LRR
STARD6	START	F25965	
FSIP1	CC (×2)	LOC348262	
		C8orf33 (FLJ20989)	
Receptor		ANKRD41 (FLJ39369)	Ankyrin domains (ANK)
NTNG2	LamNT, EGF_Lam	CCDC38 (FLJ40089)	Coiled-coil
GPR144	PTX, GPS, 7transmembrane_2	C2orf33 (GL004)	DUF800
PDCL	Similar to phosducin	NLF1 (LOC145741)	NF-κB target gene
		BU077088 (LOC284409)	Transmembrane
Ras family		BX103302 (LOC284931)	
FGD4	FYVE, RhoGEF and PH-domain-containing 4	LOC340109	
FLJ32810	Rho_GAP, SH3	C21orf111 (LOC388830)	
SIPA1L2	Rap_GAP, PDZ_signalling	LOC400236	
RAB9A	RAB	LOC400861	
SYT13	Rab effector; C2	LOC55924	
		LOC56181	DUF729
		MGC10701	
		MGC15634	
		LOC56390 (LOC388497)	
		LOC257396	

and 1 in 3×10^9 (γ-TuRC, $P \leq 0.000000003$), respectively, as calculated by hypergeometric distribution analysis.

A surprising observation was the enrichment of genes, the expression of which is normally restricted to the testis. Four of these are known to encode tumour antigens that are markedly upregulated in many tumour types including NSCLC, breast cancer and melanoma (Fig. 1e; $P \leq 0.003$ (hypergeometric distribution)). The restricted expression pattern and immunogenicity of cancer/testis antigens (CT antigens) has driven forward efforts for their use in cancer vaccines even in the absence of functional information[12]. Their identification in this screen suggests the obligatory participation of some CT antigens in aberrant cancer-cell regulatory programmes.

To probe the extent of chemosensitization that can be conferred by target gene depletion, a panel representing six functionally diverse groups from the 'high-confidence' hit list was selected. This panel included the following: CT-antigen ACRBP; the proteasome subunit PSMA6; the γ-TuRC protein TUBGCP2; a heparin sulphate transferase, HS6ST2, with significantly enriched expression in lung tumour tissue compared with normal lung (L.G. and J.D.M., unpublished observations); a vacuolar ATPase subunit, ATP6V0D2, expressed from a locus amplified in several lung cancer lines; and FGD4, a CDC42 activator. As controls for 'off-target' siRNA phenomena, we verified that each siRNA pool resulted in target gene knockdown, that at least two single siRNAs would recapitulate the phenotype

when tried separately, and that distinct pools of four more independent siRNAs against each gene also resulted in target knockdown and paclitaxel sensitization (Supplementary Fig. 2a–d).

We next examined the consequence of target depletion on responses to a broad range of concentrations of paclitaxel, vinorelbine and gemcitabine. Paclitaxel and vinorelbine impair mitotic spindle assembly through independent mechanisms that suppress microtubule dynamics[11]. In contrast, gemcitabine induces replication-stress-dependent apoptosis through DNA chain termination[13]. Several targets displayed significant collaboration with paclitaxel concentrations 1,000-fold below that used for the primary screen (Fig. 2a). Exposure to paclitaxel for a further 24–48 h magnified these differences: some groups responded to paclitaxel concentrations 10,000-fold lower than otherwise required (Fig. 2d). Cell survival curves and colony assays suggest that the decrease in cell number is a consequence of cell death rather than a transient delay in proliferation (Supplementary Fig. 3a, b). Some targets also significantly enhanced sensitivity to vinorelbine but, for the most part, only at concentrations at which vinorelbine alone detectably impaired cell viability (Fig. 2b). In contrast, target depletion had no remarkable consequence on the maximally effective concentration of gemcitabine in H1155 cells, although the non-responding cell population was decreased or eliminated in all cases (Fig. 2c).

The apparent synergy that we observed between target depletion and paclitaxel, in comparison with other chemotherapeutic agents, suggests that genome-wide chemosensitizer screens return molecular components closely related to the mode of action of a particular drug. To test this directly, we examined the consequence of target depletion on the morphology of the mitotic spindle[14]. Depletion of FGD4 resulted in a significant accumulation of otherwise normal-appearing mitotic figures in the absence of paclitaxel, indicating that this protein may be required for support of mitotic progression (Supplementary Fig. 4). Depletion of ACRBP and TUBGCP2, although not detectably affecting mitosis in the absence of drug, resulted in a marked accumulation of multipolar spindles in the presence of 10 nM paclitaxel (Fig. 3a, and Supplementary Fig. 5b). Multipolar spindle accumulation is typical after exposure to higher doses of paclitaxel in H1155 cells as well as in other cancer cell types[9,11] (data not shown). Simultaneous depletion of MAD2, an obligate component of the spindle assembly checkpoint[14], reversed the accumulation of mitotic figures with the concomitant appearance

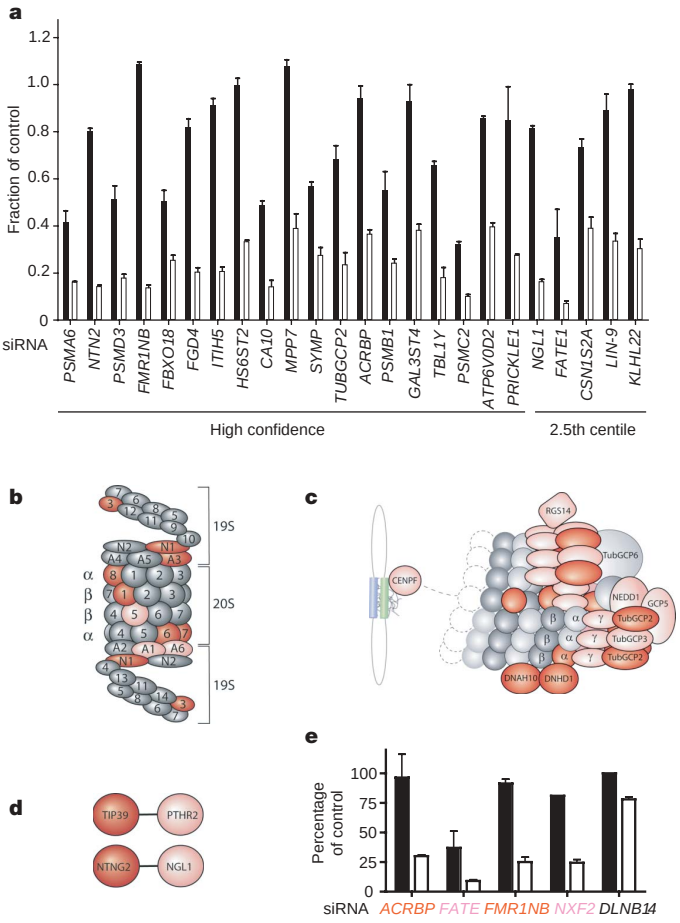


Figure 1 | Functional relationships among candidate paclitaxel-sensitizing siRNA targets. a, Retests of a panel of independently synthesized siRNA pools targeting candidate genes that modulate paclitaxel sensitivity. Results are cell viability normalized to control siRNA-transfected samples and are shown as means and s.e.m. for $n = 3$. Black bars, no paclitaxel; white bars, 10 nM paclitaxel. **b**, Proteasome. Red shading indicates satisfaction of 5% FDR, pink shading indicates satisfaction of 10% FDR. **c**, γ-TuRC and related components of the mitotic spindle apparatus. Shading is as in **b**. **d**, Ligand–receptor pairs. Shading is as in **b**. **e**, CT antigens. Results are viability normalized as a percentage of control siRNA transfected samples (*DLNB14*) and are shown as means and s.e.m. Black bars, 0 nM paclitaxel; white bars, 10 nM paclitaxel. siRNA gene shading is as in **b**. Values are representative of a minimum of three independent experiments.

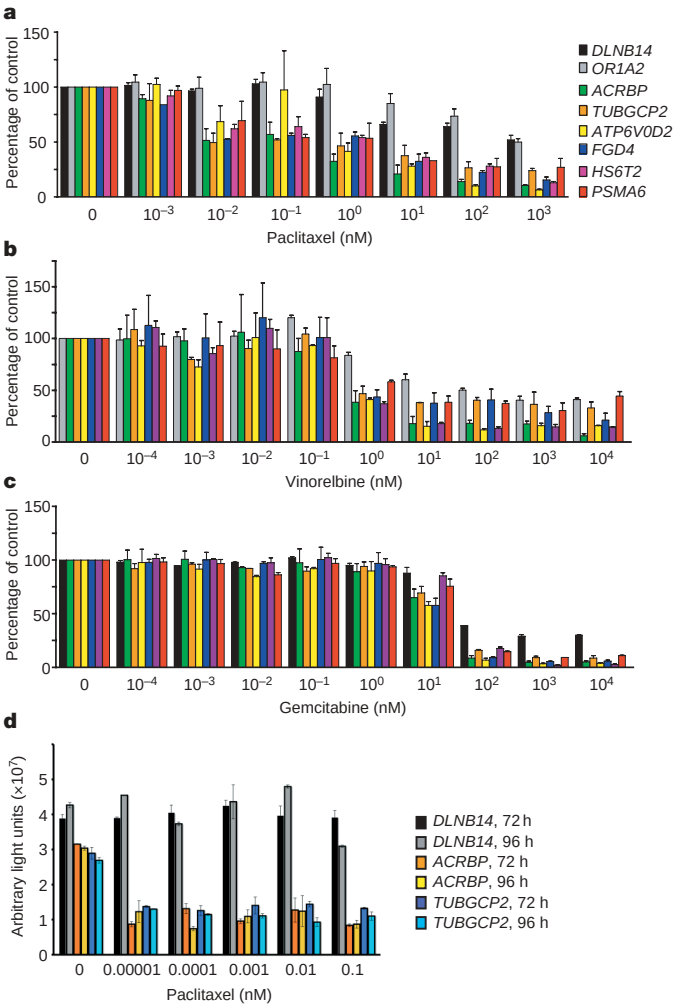


Figure 2 | Drug sensitivity profiles. a–c, H1155 transfected with siRNAs targeting the indicated genes (*DLNB1* and *OR1A2* are control siRNAs) were exposed to paclitaxel (**a**), vinorelbine (**b**) or gemcitabine (**c**) 48 h after transfection at the indicated doses for 48 h. Results are viability normalized to siRNA-transfected samples in the absence of drug and are shown as means and s.e.m. Values are representative of three independent experiments. **d**, H1155 transfected with siRNAs targeting the indicated genes were treated with paclitaxel 48 h after transfection for the indicated times. Bars are cell viability obtained with Cell Titer Glo and are shown as means and s.e.m.

of numerous micronucleated cells, indicating mitotic slippage through a defective spindle assembly checkpoint (Fig. 3a). Depletion of PSMA6, HS6ST2 and ATP6V0D2 did not affect mitotic spindle assembly (data not shown).

Given the significant genetic heterogeneity between cancer cell lines we next examined the effect of target depletion on a panel of lung lines, with diverse paclitaxel IC_{50} values, that included the NSCLC line HCC4017 and normal, non-malignant bronchial epithelial line HBEC30 (ref. 15), isolated from the same individual. Out of 12 targets tested with the patient-matched tumour and normal lines, the depletion of 4 targets selectively sensitized the tumour-derived line to low-dose paclitaxel (Supplementary Fig. 5a). Two of the four sensitizers were in the CT-antigen family. Three out of four CT antigens tested also sensitized at least one additional NSCLC line to low-dose paclitaxel with no measurable consequences on the viability of HBEC30 cells. Not surprisingly, proteasome subunit depletion was broadly effective in tumour cells in comparison with normal cells (Supplementary Fig. 5a).

We also examined the effect of ACRBP or TubGCP2 depletion on mitotic progression in these lines. Although neither ACRBP nor TUBGCP2 depletion affected cell viability as assessed by ATP concentration in H1299 and H2126 NSCLC cells, depletion of these targets did enhance paclitaxel-induced mitotic arrest (Supplementary Fig. 5b). The lack of change in viability may reflect differences in the coupling of spindle assembly checkpoint machinery to apoptosis in different cancer cell lines[16]. Consistent with this was our observation that depleting ACRBP or TUBGCP2 sensitized H1155 cells to paclitaxel-induced caspase activation (Supplementary Fig. 5d), whereas in H2126 cells, depletion of ACRBP collaborated with paclitaxel to inhibit proliferation (Supplementary Fig. 6a). In addition, the depletion of either ACRBP or TUBGCP2 in lung-tumour-derived cell lines lacking a robust spindle assembly checkpoint (HCC366, HCC15 or HCC4017) was sufficient to induce the accumulation of non-proliferating micronucleated cells, which are normally observed after exposure to paclitaxel (Fig. 3c, and Supplementary Fig. 6b, c). These observations highlight the emerging concept that products of anomalous gene-expression programmes can become engaged to buttress the fundamental biological systems required for the proliferative fitness of cancer cells. In a specific sense, aberrant expression of proteins such as ACRBP may contribute to mitotic progression in cancer cells by enhancing the robustness of an otherwise weakened mitotic spindle apparatus.

An expected outcome of genomic chemosensitizer screens is the identification of gene products that are targets of currently available compounds, indicating novel combinatorial therapeutic regimens. Our isolation of the proteasome exemplifies this relationship because collaboration between bortezomib, a proteasome inhibitor, and paclitaxel has been demonstrated clinically[17]. Isolation of ATP6V0D2, a subunit of the vacuolar ATPase (V-ATPase)[18] (Table 1 and Fig. 1a), predicts that lysosomal ATPase-inhibitors may be effective cytocidal agents in combination with paclitaxel. Salicylihalamide A was originally identified as an anti-tumour agent and was subsequently found to target V-ATPase activity directly[19,20]. Exposure of H1155 cells to increasing doses of a synthetic salicylihalamide derivative[21], RTA 203, together with low-dose paclitaxel revealed a significant collaborative impact on viability at doses well below that required for the activity of a single agent (Fig. 3d, e). This observation highlights the strong predictive power of genome-wide synthetic-lethal screens for identification of productive drug–drug interactions. We have used a high-throughput functional-genomics screening platform, together with an objective 'hit' selection criterion derived from probabilistic judgments of error rates, to produce an unbiased and high-confidence collection of the molecular components modulating chemosensitivity in lung cancer cells. The results reveal major fulcrums of the autonomous response of cancer cells to abrogation of microtubule dynamics; the results also identify therapeutic targets for combinatorial chemotherapy and highlight a major contribution of cancer-associated anomalous gene expression patterns for support of mitotic progression in cancer cells.

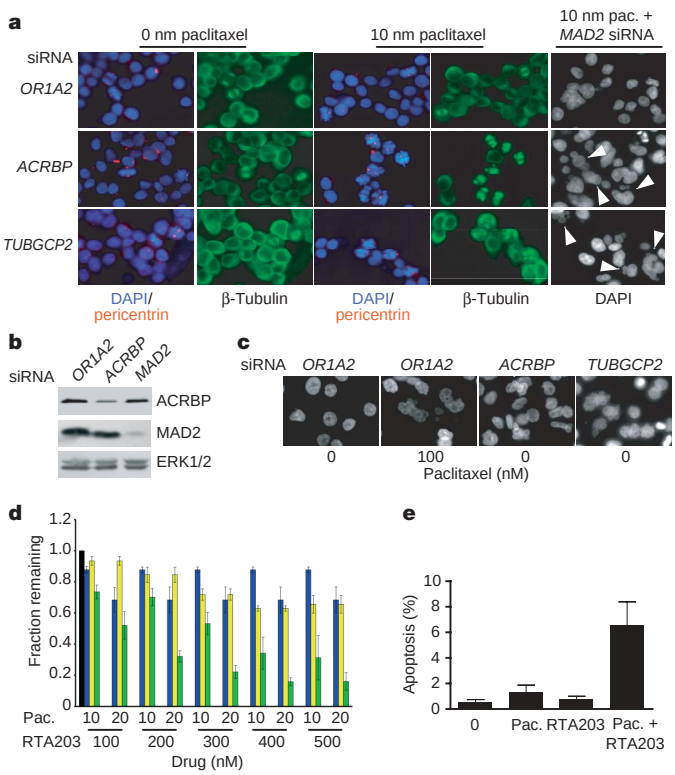


Figure 3 | Convergence of paclitaxel and sensitizer gene function on mitotic spindle integrity. **a**, At 48 h after transfection with the indicated siRNAs, H1155 cells were exposed to the indicated paclitaxel concentrations for 24 h. Microtubules, genomic DNA and centrosomes were revealed by immunostaining with β-tubulin, 4,6-diamidino-2-phenylindole (DAPI) and α-pericentrin, respectively. Arrowheads indicate the formation of micronuclei as a consequence of bypass of MAD2-dependent mitotic spindle checkpoint arrest. Pictures are representative of a minimum of five independent experiments. Pac., paclitaxel. **b,** siRNA-dependent depletion of ACRBP and MAD2 was verified by immunoblots of whole-cell lysates from **a**. **c**, The interphase nuclear morphology of HCC366 cells transfected with the indicated siRNAs was examined with DAPI. Arrowheads indicate cells containing multiple micronuclei. Pictures are representative of a minimum of three independent experiments. **d**, Collaborative impact of paclitaxel and RTA-203 on H1155 cell viability. Results are viability normalized to untreated control samples (black bar) and are shown as means and s.e.m. Yellow bars, RTA-203 alone; blue bars, paclitaxel alone; green bars, RTA-203 plus paclitaxel at the indicated doses. Values are representative of three independent experiments. **e**, Percentage apoptosis as indicated by cleaved caspase-3 immunostaining of H1155 cells after a 24-h exposure to 10 nM paclitaxel or 200 nM RTA-203, or a combination of both. Error bars show s.e.m. for four independent experiments.

METHODS

All cells were reverse-transfected with siRNA pools complexed with DharmaFECT reagent (optimized for each cell type). Cells were treated 48 h after transfection and viability was assessed after an additional 48 h. For screening data analysis, each siRNA pool was assigned a viability ratio. Viability ratios were ranked by reproducibility between three replicates for each condition, using a two-sample *t*-test followed by a Benjamini–Hochberg correction. Immunofluorescence was performed with the use of standard fixation and permeabilization protocols. Cells were stained with monoclonal β-tubulin antibodies, polyclonal pericentrin antibodies, bromodeoxyuridine or antibodies against cleaved caspase-3 followed by secondary labelling with secondary fluorescein isothiocyanate-conjugated anti-mouse antibodies or tetramethylrhodamine β-isothiocyanate-conjugated anti-rabbit antibodies. Cells were observed under a Zeiss Axioplan 2E microscope equipped with a Hamamatsu monochrome

digital black-and-white camera and Open Lab Software. Quantitative polymerase chain reaction was performed on RNA extracted from all cells with the Roche LightCycler System or the 7900HT Fast Real-Time PCR System with primers flanking at least two siRNA target sequences and lying on separate exons. Growth inhibition assays were performed with a Sulphorhodamine B protocol on cells treated for 48 hours with the indicated drugs[22]. For colony formation assays, transfected and treated cells were replated and stained with Geimsa 5 days later. Immunoblots were performed on whole-cell lysates from H1155 cells with the use of standard protocols.

A detailed description of the screening strategy and statistical analysis is given in Supplementary Fig. 1 and Supplementary Methods. Optimized transfection protocols and growth conditions for the multiple cell lines used in this study are described in Supplementary Table 6. siRNA sequences and reverse transcriptase-mediated polymerase chain reaction primers are described in Supplementary Tables 5 and 7. Methods for standard viability assays and quantification of mitotic progression are also included in Supplementary Methods.

Received 26 October 2006; accepted 20 February 2007.

1. Green, D. R. & Evan, G. I. A matter of life and death. *Cancer Cell* **1,** 19–30 (2002).
2. Chien, Y. & White, M. A. RAL GTPases are linchpin modulators of human tumour-cell proliferation and survival. *EMBO Rep.* **4,** 800–806 (2003).
3. Matheny, S. A. *et al.* Ras regulates assembly of mitogenic signalling complexes through the effector protein IMP. *Nature* **427,** 256–260 (2004).
4. Malo, N., Hanley, J. A., Cerquozzi, S., Pelletier, J. & Nadon, R. Statistical practice in high-throughput screening data analysis. *Nature Biotechnol.* **24,** 167–175 (2006).
5. Benjamini, Y. H. Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* **57,** 2890300 (1995).
6. Allison, D. B., Cui, X., Page, G. P. & Sabripour, M. Microarray data analysis: from disarray to consolidation and consensus. *Nature Rev. Genet.* **7,** 55–65 (2006).
7. Wit, E. & McClure, J. Statistical adjustment of signal censoring in gene expression experiments. *Bioinformatics* **19,** 1055–1060 (2003).
8. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* **57,** 289–300 (1995).
9. Jordan, M. A. & Wilson, L. The use and action of drugs in analyzing mitosis. *Methods Cell Biol.* **61,** 267–295 (1999).
10. Moritz, M. & Agard, D. A. Gamma-tubulin complexes and microtubule nucleation. *Curr. Opin. Struct. Biol.* **11,** 174–181 (2001).
11. Jordan, M. A. & Wilson, L. Microtubules as a target for anticancer drugs. *Nature Rev. Cancer* **4,** 253–265 (2004).
12. Scanlan, M. J., Simpson, A. J. & Old, L. J. The cancer/testis genes: review, standardization, and commentary. *Cancer Immun.* **4,** 1–15 (2004).
13. Toschi, L., Finocchiaro, G., Bartolini, S., Gioia, V. & Cappuzzo, F. Role of gemcitabine in cancer therapy. *Future Oncol.* **1,** 7–17 (2005).
14. Weaver, B. A. & Cleveland, D. W. Decoding the links between mitosis, cancer, and chemotherapy: The mitotic checkpoint, adaptation, and cell death. *Cancer Cell* **8,** 7–12 (2005).
15. Ramirez, R. D. *et al.* Immortalization of human bronchial epithelial cells in the absence of viral oncoproteins. *Cancer Res.* **64,** 9027–9034 (2004).
16. Rieder, C. L. & Maiato, H. Stuck in division or passing through: what happens when cells cannot satisfy the spindle assembly checkpoint. *Dev. Cell* **7,** 637–651 (2004).
17. Davies, A. M. *et al.* Bortezomib-based combinations in the treatment of non-small-cell lung cancer. *Clin. Lung Cancer* **7** (Suppl 2), S59–S63 (2005).
18. Kawasaki-Nishi, S., Nishi, T. & Forgac, M. Proton translocation driven by ATP hydrolysis in V-ATPases. *FEBS Lett.* **545,** 76–85 (2003).
19. Boyd, M. R. *et al.* Discovery of a novel antitumor benzolactone enamide class that selectively inhibits mammalian vacuolar-type (H^+)-ATPases. *J. Pharmacol. Exp. Ther.* **297,** 114–120 (2001).
20. Xie, X. S. *et al.* Salicylihalamide A inhibits the V0 sector of the V-ATPase through a mechanism distinct from bafilomycin A1. *J. Biol. Chem.* **279,** 19755–19763 (2004).
21. Wu, Y., Liao, X., Wang, R., Xie, X. S. & De Brabander, J. K. Total synthesis and initial structure–function analysis of the potent V-ATPase inhibitors salicylihalamide A and related compounds. *J. Am. Chem. Soc.* **124,** 3245–3253 (2002).
22. Skehan, P. *et al.* New colorimetric cytotoxicity assay for anticancer-drug screening. *J. Natl Cancer Inst.* **82,** 1107–1112 (1990).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the National Cancer Institute, the Robert E. Welch Foundation, the Susan G. Komen Foundation, the Department of Defense Congressionally Directed Medical Research Program and the National Cancer Institute Lung Cancer Specialized Program of Research Excellence.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.A.W. (michael.white@utsouthwestern.edu).

LETTERS

Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner

Eric A. Galburt[1]*†, Stephan W. Grill[1]*†, Anna Wiedmann[2], Lucyna Lubkowska[5], Jason Choy‡, Eva Nogales[1,2,4], Mikhail Kashlev[5] & Carlos Bustamante[1,2,3,4]

RNA polymerase II (RNAP II) is responsible for transcribing all messenger RNAs in eukaryotic cells during a highly regulated process that is conserved from yeast to human[1], and that serves as a central control point for cellular function. Here we investigate the transcription dynamics of single RNAP II molecules from *Saccharomyces cerevisiae* against force and in the presence and absence of TFIIS, a transcription elongation factor known to increase transcription through nucleosomal barriers[2]. Using a single-molecule dual-trap optical-tweezers assay combined with a novel method to enrich for active complexes, we found that the response of RNAP II to a hindering force is entirely determined by enzyme backtracking[3–6]. Surprisingly, RNAP II molecules ceased to transcribe and were unable to recover from backtracks at a force of 7.5 ± 2 pN, only one-third of the force determined for *Escherichia coli* RNAP[7,8]. We show that backtrack pause durations follow a $t^{-3/2}$ power law, implying that during backtracking RNAP II diffuses in discrete base-pair steps, and indicating that backtracks may account for most of RNAP II pauses. Significantly, addition of TFIIS rescued backtracked enzymes and allowed transcription to proceed up to a force of 16.9 ± 3.4 pN. Taken together, these results describe a regulatory mechanism of transcription elongation in eukaryotes by which transcription factors modify the mechanical performance of RNAP II, allowing it to operate against higher loads.

Promoter-based initiation of the 12-subunit RNAP II requires the complex assembly of a host of multi-component general transcription factors[1,9] and has therefore eluded researchers attempting to follow eukaryotic transcription at the single-molecule level. To overcome this difficulty, we adapted a previously described method that bypasses promoter-based initiation and assembles elongation complexes piecewise in the absence of factors[10] (Supplementary Information). Sporadic single-molecule activity was observed in this way (data not shown). To establish a robust assay, we increased the proportion of active elongation complexes by selecting for enzymes that responded to an initial nucleotide triphosphate (NTP) pulse (Fig. 1a, and Supplementary Information). Complexes that transcribed to a nucleotide starvation stall site protected an overlapping restriction site from digestion by the corresponding endonuclease. Inactive complexes failed to protect the restriction site and were digested away (Fig. 1a, and Supplementary Information). This pulse–digest method increased the yield of active elongation complexes approximately fivefold (data not shown).

After pulse–digest, a single ternary complex with 9.8 kb of template DNA was tethered between two polystyrene beads held in place by two single-beam optical traps[6,11,12] (Supplementary Information). On addition of 1 mM NTPs, RNAP II began to translocate along the DNA, shortening the tether between the two beads (Fig. 1b, and Supplementary Fig. 1). Force was monitored in both traps using laser beam deflections and the average value was converted to enzyme position along the template with the worm-like-chain model for DNA elasticity[13] (Supplementary Information). We observed continuous runs of transcription interrupted by pauses (Fig. 1c), which were scored and removed with a velocity-threshold algorithm to obtain a pause-free elongation velocity[14]. The average pause-free velocity was 12.2 ± 4.5 nucleotides per second ($N = 33$, mean ± s.d. unless otherwise noted; Fig. 2b, Supplementary Fig. 7 and Supplementary Table 1), which is comparable to bulk data from RNAP II[15] and to single-molecule data from the bacterial enzyme[7,11,16]. An experiment ended when, having ceased to transcribe, no resumption of activity was observed for 10 min. These experiments revealed that RNAP II was able to transcribe up to forces of 7.5 ± 2.0 pN (Fig. 2a, and Supplementary Table 1). Interestingly, although the eukaryotic enzyme translocates at rates comparable to those of its prokaryotic counterpart, it can only transcribe against a force about one-third that of the bacterial enzyme[8,16,17].

Next, we looked at the relationship between pause-free velocity and force[8,16]. Because velocity differed significantly between enzymes[8,16,17] (Fig. 2b), we preserved the force–velocity (F–V) relationship of individual enzymes by normalizing both variables[8]. Pause-free velocity shows no force dependence up to the average force of 7.5 pN (Fig. 2c), indicating that under physiological and saturating NTP concentrations, translocation is not rate limiting at these forces (also true at 120 μM NTP, data not shown). At 7.5 pN, transcription rates drop sharply to zero in contrast to the gradual decrease observed for the bacterial enzyme[8,11]. We then asked whether this sharp decrease is due to enzyme translocation becoming rate limiting at 7.5 pN or if it reflects the force sensitivity of another process. A single-parameter fit of the F–V relationship to a generalized Boltzmann scheme[8] yielded an unphysical distance to the translocation transition state of 152 ± 8 nucleotides. This suggests that the force dependence of a process other than NTP-dependent forward translocation determines the behaviour of the eukaryotic enzyme in this force regime. Thus, 7.5 pN is not the maximum force that RNAP II can generate (that is, its thermodynamic stall force), but instead represents an 'operational force limit'.

What is the physical origin of the strong force sensitivity observed around 7.5 pN? Data obtained at the maximum spatial resolution allowed by our experiments (~3 nucleotides) revealed that arrest strongly correlates with a lengthening of the tether, an observation consistent with backtracking (Fig. 1c, inset). Backtracking is independent of NTP hydrolysis and involves a process wherein the enzyme moves upstream on the DNA template while the RNA–DNA register is maintained. This process results in an inactive state

[1]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [2]Department of Molecular and Cell Biology, [3]Department of Chemistry and [4]Howard Hughes Medical Institute, University of California, Berkeley 94720, USA. [5]NCI Center for Cancer Research, Frederick, Maryland 21702, USA. †Present address: Max Planck Institute for the Physics of Complex Systems, 01187 Dresden, Germany. ‡Deceased.
*These authors contributed equally to this work.

because the 3′-end of the transcript is no longer at the active site[3,4,18] (Supplementary Fig. 2). Fifty per cent of all the experiments ended with a terminal backtrack, with an average length of 7.2 ± 3.0 nucleotides. We note that the real average terminal backtrack distance must be considerably smaller (we estimate ~3 nucleotides) because we expect to miss a significant fraction of backtrack events less than 3 nucleotides in size (Supplementary Fig. 3). These observations strongly implicate backtracking as the origin of the force sensitivity of the enzyme at 7.5 pN.

Because the backtracked state is physically displaced along the template, an applied force should affect either the rate of entry to or the rate of exit from this backtracked state, or both. Thus, we sought to characterize the force dependence of these rates. By looking at the effects of assisting force on arrest, studies on the bacterial polymerase have inferred that force primarily affects the ability of the enzyme to recover from backtracks[16]. To determine the force dependence of backtrack return for the eukaryotic enzyme, we measured the fraction of observed backtracks that did not recover within 10 min in different force ranges (Fig. 3a, filled red circles). The return probability is highly sensitive to force: above 8 pN, less than 50% of the enzymes are able to recover from the backtracked state. To determine the force dependence of backtrack entry, we measured the probability of observing a backtrack event within one second of active transcription. Below the operational force limit of the enzyme, this probability is 0.02 ± 0.012 and exhibits no force dependence (Fig. 3a, filled blue circles). To conclude, force does not significantly modulate the probability of entering a backtracked state in the operational force range of the enzyme. The probability of returning from the backtracked state is, however, highly force sensitive and the response of RNAP II to force in the absence of factors is entirely dominated by this effect.

We have shown that RNAP II can only transcribe against a force about one-third that of its bacterial counterpart as it can no longer return from the backtracked state. The maximum force that RNAP II can generate during translocation, however, should be higher than this operational force limit. To measure this thermodynamic stall force, we devised an experiment that would allow us to observe transcription at forces beyond the operational force limit of RNAP II. To this end, we performed 'force jump' experiments during active enzyme translocation[19]. In these experiments, the force on the enzyme was rapidly increased by suddenly displacing one of the two traps (Supplementary Fig. 4). After one second, the original position of the trap was restored, and the velocity of the enzyme during the jump was measured. In these experiments, enzymes were seen to transcribe beyond the operational force limit determined above. Transcription was observed for 50% (6/12) of jumps between 14 to 20 pN, for 17% (1/6) of jumps between 20 to 25 pN, and 0% (0/12) of

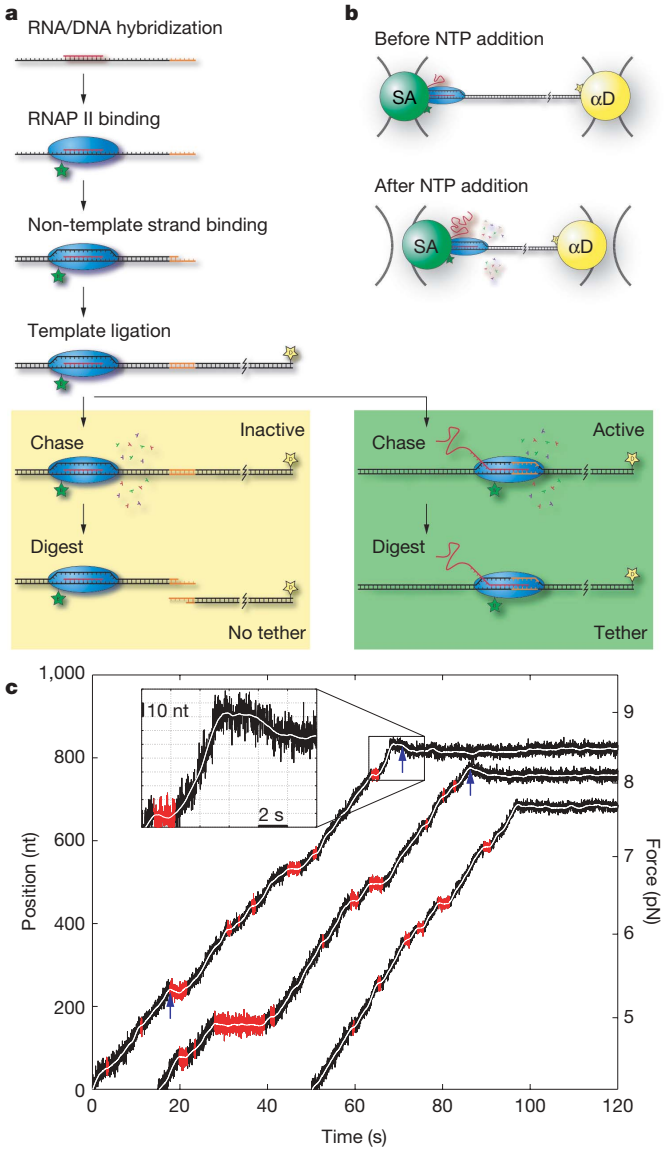


Figure 1 | Single-molecule transcription. **a**, Elongation complexes are formed as previously described[10] (Supplementary Methods; biotinylated RNAP II, blue with green star; digoxigenin, yellow star) and ligated to the transcription template. After pulse–digest, inactive complexes (yellow shading) are digested away and only active complexes (green shading) can form tethers. **b**, Passive mode dual-trap optical tweezer. After adding NTPs, transcription begins, the tether shortens and the load increases (Supplementary Fig. 1). **c**, Template position and force versus time (100 Hz bandwith, 3rd order Savitzky–Golay filter with time constant of 2.5 s). Enzymes exhibit elongation (black), pausing (red), backtracking (blue arrows and inset) and arrest. Nucleotides, nt.

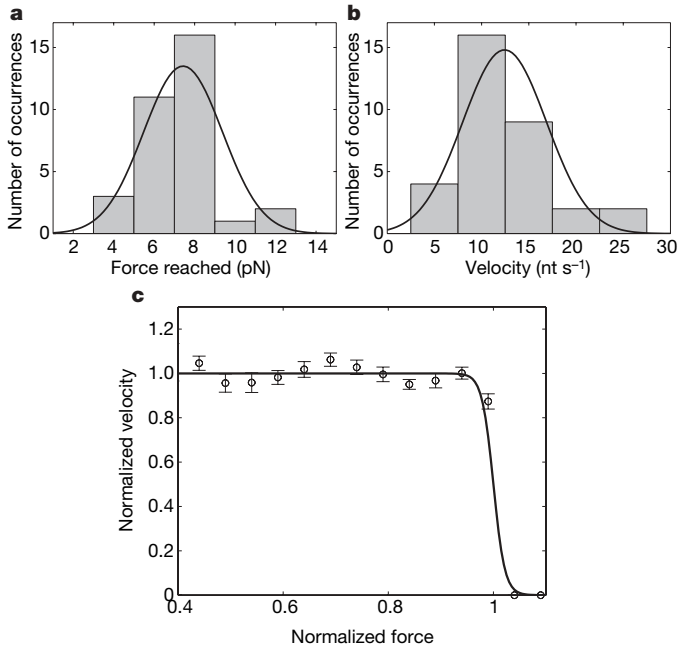


Figure 2 | Force–velocity analysis. **a**, A gaussian fit of the histogram of forces reached shows that RNAP II ceases transcription at a force of 7.5 ± 2.0 pN (mean ± s.d., $R^2 = 0.82$, $N = 33$). **b**, A gaussian fit of the pause-free transcription velocity histogram yields an average of 12.2 ± 4.5 nt s^{-1} (mean ± s.d., $R^2 = 0.85$, $N = 33$). **c**, Normalized force–velocity curve (Supplementary Information). Velocity is uncorrelated with force (error bars, s.e.m.) for normalized forces smaller than 1 ($r^2 = 7 \times 10^{-3}$). The solid line represents a fit to a generalized Boltzmann scheme[8] (single parameter fit, $R^2 = 0.97$).

jumps over 25 pN. Enzymes that continued to transcribe in these experiments displayed pause-free velocities not significantly different from velocities measured at lower forces (data not shown), indicating that translocation is still not rate limiting at these increased forces. These results indicate that the thermodynamic stall force of RNAP II is higher than 20 pN at physiological NTP concentrations. Significantly, measurements of transcriptional velocity beyond ~25 pN were foiled because the probability of backtracking within one second increased dramatically (Fig. 3a, filled blue squares): 17/21 enzymes backtracked during the jump at jump forces beyond 20 pN, and 100% (7/7) of enzymes backtracked above 30 pN. In conclusion, both backtrack entry and exits are sensitive to force, however, the former only at very high forces, whereas the latter determines the operational force limit of the enzyme.

We have shown that the main response of RNAP II to force is backtracking, which causes the enzyme to pause. Next we ask if it is possible that all transcriptional pauses in RNAP II are due to backtracking. However, with a resolution of 3 to 4 nucleotides at physiological NTP concentrations[11], it is not possible to determine whether or not brief pauses are associated with backtracking events. Specifically, backtrack displacements of a couple of nucleotides will not be recognized (Supplementary Fig. 3 and Supplementary Information). Because it is easier to determine the duration of a pause than its backtracking distance, we identified enzyme pauses with a computer algorithm and analysed the distribution of pause lifetimes[14]. Analysis of the pause time durations (t) revealed that the distribution of pause lifetimes follows a $t^{-3/2}$ power law (Fig. 3c, and Supplementary Information). This observation implies a mechanism in which, during a pause, a polymerase diffuses among many intermediate states and ends the pause when it diffusively realigns the dislocated 3′-end of the transcript with the active site to resume elongation. Significantly, all pause durations, short and long, follow this power law distribution, implying that the molecular mechanism underlying most transcriptional pauses in RNAP II is the same, that is, diffusive backtracking. Note that the distribution deviates from the $t^{-3/2}$ power law only for pauses longer than 10 s, as is to be expected from the effect of the opposing force on the polymerase (manuscript in preparation).

Because the operational force limit of the enzyme is determined by the force dependence of exit from backtracked states, it follows that the enzyme should be able to transcribe to higher forces in conditions that increase its probability to return from these states. Specifically, the polymerase would be expected to transcribe up to forces in the 20 pN range, where the probability of backtracking would limit transcriptional progress (Fig. 3a). To test this prediction, we performed experiments in the presence of the transcription factor TFIIS. This factor is involved in both the initiation and elongation of RNAP II[20], is functionally homologous to the Gre factors in bacteria[21], and may have an important role in RNA proofreading[22]. TFIIS binds RNAP II, catalyses the enzyme's intrinsic RNA cleavage activity, and rescues backtracked states by producing a new 3′-end that is aligned with the enzyme's active site[23,24] (Supplementary Fig. 2).

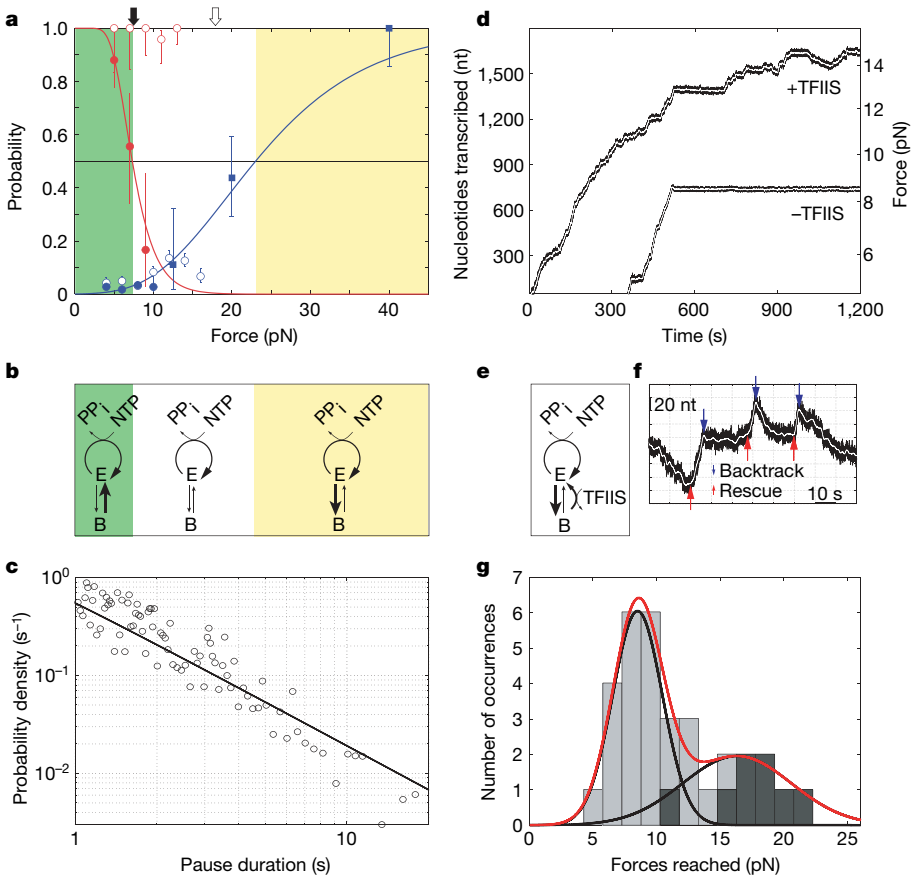


Figure 3 | Backtrack entry and exit. **a**, The probability of entering a backtrack within one second of transcription (blue markers) and the probability of returning from a backtrack within 10 min (red markers) versus force. Data are shown in the absence (filled markers) and presence (open markers) of TFIIS, and arrows indicate the respective forces reached. Square markers represent data from force jump experiments. Error bars represent s.e.m. and solid lines indicate trends. **b**, Illustrated kinetic schemes for three regions of force where return to elongation (E) from a backtrack (B) is likely (green), unlikely (white and yellow) and where backtrack entry is likely (yellow). **c**, Distribution of measured pause durations with a single-parameter fit of a $t^{-3/2}$ power-law ($R^2 = 0.84$). **d**, Template position and force versus time in the presence and absence of TFIIS (Supplementary Fig. 5). **e**, Illustrated scheme in the presence of TFIIS. **f**, An example of cycles of backtracking and TFIIS rescue at 18 pN. **g**, Histogram of forces reached in the presence of TFIIS. Instances of cycling between backtracking and rescue are indicated in dark grey. Solid lines indicate a double gaussian fit with means of 8.5 ± 2.0 and 16.5 ± 3.5 pN (mean ± s.d., $R^2 = 0.96$, $N = 32$).

TFIIS was expressed and purified as described[23] and introduced at a saturating concentration (600 nM), concomitant with NTPs. Seventy five per cent of the runs obtained in the presence of TFIIS (24/32) were not significantly different from runs obtained in the absence of TFIIS (Fig. 3g, and Supplementary Information). However, 25% of the transcription runs (8/32) exhibited a dramatic change in behaviour (Fig. 3d): instead of arresting after a terminal backtrack, these enzymes never ceased to transcribe but repeatedly switched between backtracking and active transcription (Fig. 3d, f, and Supplementary Fig. 5) and translocated to significantly higher forces (16.9 ± 3.4 pN, Fig. 3d, g). This behaviour was never observed in the absence of TFIIS (0/33). At this new operational force limit, the average distance backtracked before rescue equals the average distance transcribed before entering a backtracked state (Fig. 3d, f, g). Again, pause-free velocity was unchanged even at the increased opposing forces accessible in the presence of TFIIS (Supplementary Fig. 6). To conclude, TFIIS increases the operational force limit of RNAP II more than twofold by accelerating backtrack exit and allowing the polymerase to proceed to forces where the rate of backtrack entry is significantly increased (Fig. 3a). These results rationalize observations that TFIIS promotes transcription through transcriptional blocks such as the *lac* operon repressor[25] and the nucleosome *in vitro*[2]. Additionally, they predict that bacterial enzymes might transcribe to higher forces in GreB's presence.

The low intrinsic operational force limit of RNAP II, coupled with the dramatic effect of TFIIS, indicates the possibility of transcriptional regulation through a TFIIS-dependent switch of the polymerase's operational force limit. Although it does not seem that TFIIS levels are regulated directly in yeast (C. M. Kane, personal communication), the ability of TFIIS to bind the polymerase is regulated by additional factors in multi-cellular eukaryotes[26]. The results described here set the stage for the investigation of eukaryotic transcription elongation through nucleosomal arrays in the presence and absence of transcription elongation factors, using single-molecule manipulation methods.

METHODS

A detailed description of materials and methods is given in Supplementary Information.

Transcription initiation. Elongation complexes were formed through the ordered addition of a 54 nucleotide template DNA, 9 nucleotide RNA primer, biotinylated RNAP II, and complementary 50 nucleotide non-template DNA strand[10,27]. Elongation complexes were then ligated to a 9.8 kilobase double-stranded DNA template with a downstream digoxigenin label. Oligonucleotide sequences and other details are provided in Supplementary Information.

Optical trapping. The optical trap used for these experiments is based on a setup previously described[28]. To reduce drift, the dual-beam, single-trap setup was converted to a single-beam, dual-trap machine by slightly overfilling each objective. The experimental setup is described in more detail in the Supplementary Information.

Backtrack detection. Backtracks were detected by looking at the first 4 s of a pause and asking whether the enzyme moved backwards by more than 3 nucleotides. These strict limits ensured that backtrack signals were not generated by noise.

Received 11 November 2006; accepted 23 February 2007.
Published online 14 March 2007.

1. Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21,** 327–335 (1996).
2. Kireeva, M. L. *et al.* Nature of the nucleosomal barrier to RNA polymerase II. *Mol. Cell* **18,** 97–108 (2005).
3. Komissarova, N. & Kashlev, M. Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3′ end of the RNA intact and extruded. *Proc. Natl Acad. Sci. USA* **94,** 1755–1760 (1997).
4. Komissarova, N. & Kashlev, M. RNA polymerase switches between inactivated and activated states by translocating back and forth along the DNA and the RNA. *J. Biol. Chem.* **272,** 15329–15338 (1997).
5. Nudler, E., Mustaev, A., Lukhtanov, E. & Goldfarb, A. The RNA–DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* **89,** 33–41 (1997).
6. Shaevitz, J. W., Abbondanzieri, E. A., Landick, R. & Block, S. M. Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature* **426,** 684–687 (2003).
7. Davenport, R. J., Wuite, G. J., Landick, R. & Bustamante, C. Single-molecule study of transcriptional pausing and arrest by *E. coli* RNA polymerase. *Science* **287,** 2497–2500 (2000).
8. Wang, M. D. *et al.* Force and velocity measured for single molecules of RNA polymerase. *Science* **282,** 902–907 (1998).
9. Hahn, S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nature Struct. Mol. Biol.* **11,** 394–403 (2004).
10. Komissarova, N., Kireeva, M. L., Becker, J., Sidorenkov, I. & Kashlev, M. Engineering of elongation complexes of bacterial and yeast RNA polymerases. *Methods Enzymol.* **371,** 233–251 (2003).
11. Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R. & Block, S. M. Direct observation of base-pair stepping by RNA polymerase. *Nature* **438,** 460–465 (2005).
12. Moffitt, J., Chemla, Y., Izhaky, D. & Bustamante, C. Differential detection of dual traps improves the spatial resolution of optical tweezers. *Proc. Natl Acad. Sci. USA* **103,** 9006–9011 (2006).
13. Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. Entropic elasticity of lambda-phage DNA. *Science* **265,** 1599–1600 (1994).
14. Neuman, K. C., Abbondanzieri, E. A., Landick, R., Gelles, J. & Block, S. M. Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell* **115,** 437–447 (2003).
15. Izban, M. G. & Luse, D. S. Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. *J. Biol. Chem.* **267,** 13647–13655 (1992).
16. Forde, N. R., Izhaky, D., Woodcock, G. R., Wuite, G. J. L. & Bustamante, C. Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by *Escherichia coli* RNA polymerase. *Proc. Natl Acad. Sci. USA* **99,** 11682–11687 (2002).
17. Yin, H. *et al.* Transcription against an applied force. *Science* **270,** 1653–1657 (1995).
18. Kulish, D. & Struhl, K. TFIIS enhances transcriptional elongation through an artificial arrest site *in vivo*. *Mol. Cell. Biol.* **21,** 4162–4168 (2001).
19. Li, P. T. X., Collin, D., Smith, S. B., Bustamante, C. & Tinoco, I. Probing the mechanical folding kinetics of TAR RNA by hopping, force-jump, and force-ramp methods. *Biophys. J.* **90,** 250–260 (2006).
20. Fish, R. N. & Kane, C. M. Promoting elongation with transcript cleavage stimulatory factors. *Biochim. Biophys. Acta* **1577,** 287–307 (2002).
21. Borukhov, S., Lee, J. & Laptenko, O. Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Mol. Microbiol.* **55,** 1315–1324 (2005).
22. Jeon, C. & Agarwal, K. Fidelity of RNA polymerase II transcription controlled by elongation factor TFIIS. *Proc. Natl Acad. Sci. USA* **93,** 13677–13682 (1996).
23. Awrey, D. E. *et al.* Yeast transcript elongation factor (TFIIS), structure and function. II: RNA polymerase binding, transcript cleavage, and read-through. *J. Biol. Chem.* **273,** 22595–22605 (1998).
24. Weilbaecher, R. G., Awrey, D. E., Edwards, A. M. & Kane, C. M. Intrinsic transcript cleavage in yeast RNA polymerase II elongation complexes. *J. Biol. Chem.* **278,** 24189–24199 (2003).
25. Reines, D. & Mote, J. Jr. Elongation factor SII-dependent transcription by RNA polymerase II through a sequence-specific DNA-binding protein. *Proc. Natl Acad. Sci. USA* **90,** 1917–1921 (1993).
26. Palangat, M., Renner, D. B., Price, D. H. & Landick, R. A negative elongation factor for human RNA polymerase II inhibits the anti-arrest transcript-cleavage factor TFIIS. *Proc. Natl Acad. Sci. USA* **102,** 15036–15041 (2005).
27. Kireeva, M. L., Lubkowska, L., Komissarova, N. & Kashlev, M. Assays and affinity purification of biotinylated and nonbiotinylated forms of double-tagged core RNA polymerase II from *Saccharomyces cerevisiae*. *Methods Enzymol.* **370,** 138–155 (2003).
28. Smith, S. B., Cui, Y. & Bustamante, C. Optical-trap force transducer that operates by direct measurement of light momentum. *Methods Enzymol.* **361,** 134–162 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. R. Chemla, W. Cheng, M. Cruse, S. Dumont, N. R. Forde, B. Ibarra, D. Izhaky, C. Kane, S. Kostek, J. Moffit, J. M. R. Parrondo, M. Peris, S. Plyasunov, A. Ruiz and S. B. Smith for experimental assistance and helpful discussions. This research was supported by a DOE grant. E.A.G. was supported by a Jane Coffin Childs Postdoctoral Fellowship. S.W.G. was supported first by an EMBO Long Term Fellowship and subsequently by a Helen Hay Whitney Postdoctoral Fellowship. We dedicate this manuscript to Jason Choy.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.B. (carlos@alice.berkeley.edu).

CORRIGENDUM

doi:10.1038/nature05710

The Polycomb group protein EZH2 directly controls DNA methylation

Emmanuelle Viré, Carmen Brenner, Rachel Deplus, Loïc Blanchon, Mario Fraga, Céline Didelot, Lluis Morey, Aleyde Van Eynde, David Bernard, Jean-Marie Vanderwinden, Mathieu Bollen, Manel Esteller, Luciano Di Croce, Yvan de Launoit & François Fuks

Nature **439**, 871–874 (2006)

It has been drawn to our attention that in this Letter some of the lanes in Figures 2a, 2c, 4a and Supplementary Fig. S1 were inadvertently missing. A description of how the figures were made is provided in Supplementary Information to this Corrigendum, together with the corrected figure panels. Our results and conclusions are unaffected by these oversights.

Supplementary Information is linked to the online version of the Corrigendum at www.nature.com/nature.

ADDENDUM

doi:10.1038/nature05727

Molecular basis for interaction of the protein tyrosine kinase ZAP-70 with the T-cell receptor

Marcos H. Hatada, Xiaode Lu, Ellen R. Laird, Jeremy Green, Jay P. Morgenstern, Meizhen Lou, Chris S. Marr, Thomas B. Phillips, Mary K. Ram, Kelly Theriault, Mark J. Zoller & Jennifer L. Karas

Nature **377**, 32–38 (1995)

The Protein Data Bank ID at the RCSB is 2OQ1 for the ZAP70 coordinates.

CORRIGENDUM

doi:10.1038/nature05710

The Polycomb group protein EZH2 directly controls DNA methylation

Emmanuelle Viré, Carmen Brenner, Rachel Deplus, Loïc Blanchon, Mario Fraga, Céline Didelot, Lluis Morey, Aleyde Van Eynde, David Bernard, Jean-Marie Vanderwinden, Mathieu Bollen, Manel Esteller, Luciano Di Croce, Yvan de Launoit & François Fuks

Nature **439**, 871–874 (2006)

It has been drawn to our attention that in this Letter some of the lanes in Figures 2a, 2c, 4a and Supplementary Fig. S1 were inadvertently missing. A description of how the figures were made is provided in Supplementary Information to this Corrigendum, together with the corrected figure panels. Our results and conclusions are unaffected by these oversights.

Supplementary Information is linked to the online version of the Corrigendum at www.nature.com/nature.

ADDENDUM

doi:10.1038/nature05727

Molecular basis for interaction of the protein tyrosine kinase ZAP-70 with the T-cell receptor

Marcos H. Hatada, Xiaode Lu, Ellen R. Laird, Jeremy Green, Jay P. Morgenstern, Meizhen Lou, Chris S. Marr, Thomas B. Phillips, Mary K. Ram, Kelly Theriault, Mark J. Zoller & Jennifer L. Karas

Nature **377**, 32–38 (1995)

The Protein Data Bank ID at the RCSB is 2OQ1 for the ZAP70 coordinates.

naturejobs

THE CAREERS MAGAZINE FOR SCIENTISTS

It may sound odd, but a recent *Nature* initiative reminds me of the film *Field of Dreams*, starring Kevin Costner. The film features a recurring phrase — "If you build it, he will come" — which is whispered to Costner's character by a disembodied voice. The voice exhorts Costner to carve out a baseball diamond on his farm's profitable cornfield. The results are not altogether predictable. Yes, deprived of a lucrative crop, his farm slides towards bankruptcy, but the makeshift baseball pitch also plays host to the ghosts of disgraced baseball players. Most importantly for Costner's character, the pitch also allows him to make peace with his deceased father. And, even more remarkably, the real-life diamond has since become a tourist attraction.

The film's famous phrase is now commonplace in the popular media, reworked as "If we build it, they will come" — and in that spirit we have seen some unexpected events at *Nature*'s Network website. When we launched it last June, we were not sure who would join in or what they would say. The website was set up as a social networking site to connect like-minded people through forums, message boards and blogs. We hoped that scientists would meet up in this virtual world to discuss their research and build new collaborations. Today the site seems to have evolved into a forum for young scientists to discuss their career aspirations and obstacles. British postdocs, researchers at the Marie Curie Research Institute in Oxted, UK, and young scientists in the Boston area have all established a presence. And many postdocs and graduate students are adding blogs or linking out to existing ones.

With the launch of the London network last month, and other international nodes to follow, we expect such activity to grow. This growth shows that young scientists are hungry for career information and are willing to share what they know. We welcome such activity, even though we don't know which direction the content will take. If you haven't visited yet, please check out this growing network at http://network.nature.com and feel free to join in the discussions or start your own, and blog away. We built it, and we're pleased you're coming.

Paul Smaglik, *Naturejobs* editor

CONTACTS
Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
***Nature*events:**
Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:
Muriel Lestringuez (4994)
UK/Ireland/Italy/RoW:
Nils Moeller (4953)
Scandinavia/Spain/Portugal:
Evelina Rubio-Morgan (4973)
Germany/Austria/The Netherlands:
Reya Silao (4970)
Online Job Postings:
Matthew Ward (+44 (0) 20 7014 4059)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:**
Tom Hancock
***Naturejobs* online production:**
Catherine Alexander

US Head Office, New York
75 Varick Street,
9th Floor,
New York,
NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

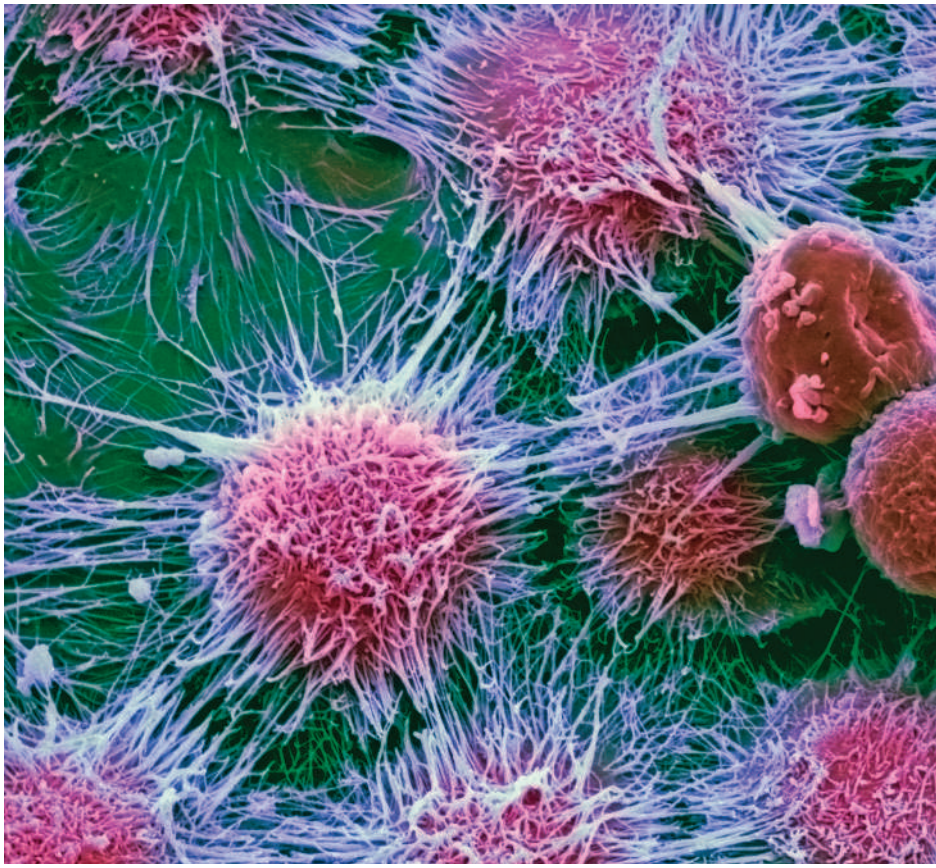
US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku,
Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com

Genes versus cancer

Just a decade has changed cancer research beyond recognition, but adaptable skills and teamwork lead to exciting possibilities. **Ricki Lewis** looks at the new face of cancer genetics.



D. MCCARTHY/SPL

Until very recently, cancer care followed a one-size-fits-all approach. Diagnosis and treatment were based on the affected body part or, for the better-studied cancers, on affected cell type. Then subtyping became popular — identifying hormone-receptor targets on tumour cells. Breast cancer, for example, was treated with a drug that blocks oestrogen receptors. Subsequent approaches focused on mutations in specific oncogenes and tumour-suppressor genes. This led to improved diagnoses and the development of spectacularly successful drugs such as Gleevec (for a form of leukaemia) and Herceptin (for a form of breast cancer).

Today, cancer genetics has mushroomed, as have related research opportunities. Two recent closer looks at mutations in common cancers, one from a group at Johns Hopkins University in Baltimore and another from the Sanger Institute in Hinxton, UK, found hundreds of implicated genes, some known but many novel. The magnitude of several translational research efforts indicates a bright future for this area of cancer research. A consortium called TRANSBIG, for example, intends to evaluate genetic signatures in cancer and has 40 partners in 21 countries. In the United States, the National Human Genome Research Institute and National Cancer Institute have invested $100 million to set up a three-year pilot programme at a dozen research centres to build a Cancer Genome Atlas of mutations. The institutes plan to expand to 60 centres by 2012.

Subtle shifts

Cancers are also increasingly being described in a different way — by subtle shifts in the expression of key genes. By quantifying messenger RNA transcription in cancer cells, researchers can chronicle changing gene expression and better detect, track and treat cancers with unprecedented specificity.

"Individualized medicine is here today. It is going to be a hugely important field, with a great impact both on research and cancer care," says Steve Shak, co-founder and chief medical officer of Genomic Health in Redwood City, California. The company's Oncotype DX breast-cancer assay translates the expression of 21 genes in tumour samples into an estimate of the risk of recurrence and probability of benefit from certain chemotherapy treatments.

Dimitri Semarov, tumour-genomics group leader at Abbott, in Abbott Park, Illinois, suggests that it's no longer a matter of studying individual oncogenes or tumour suppressors. "The biggest shift in cancer research is the adoption of the genome-wide view," he says. Researchers now have a more complete picture of the cancer genome, and they need more complex statistics to handle gigabytes of data. They are also attempting to identify other factors that affect outcome and responses, namely new biomarkers.

N. SIMON

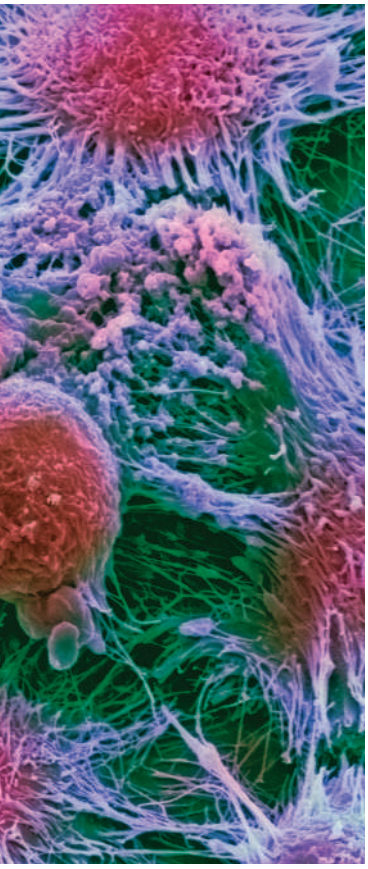
C. SCHOLL

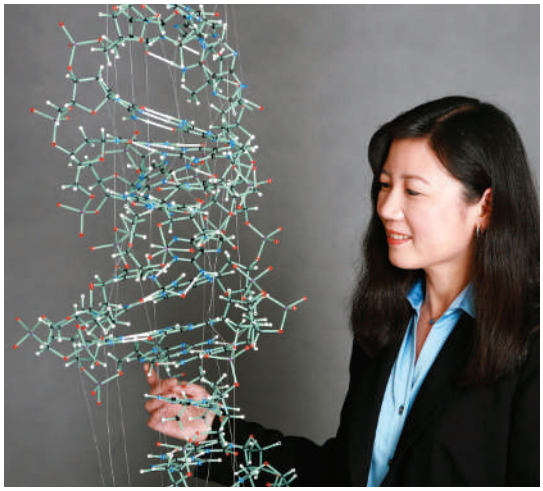
Richard Simon (top) and Stefan Fröhling.

Analysing gene-expression profiles in cancer requires skills in genomics and statistics — which usually means researchers working in teams. Genes of interest must be identified; their expression levels visualized, compared and measured with DNA microarray experiments; and then the data analysed and validated.

Development of the Oncotype DX test illustrates the research trajectory, and the diverse researcher skills for developing a test that can match patients to treatments. The first step is to identify a need. For breast cancer, this was apparent. "Chemotherapy is recommended for all, but that was because we didn't know in whom the cancer would recur, and who would benefit from chemo," says Shak. So the team consulted published studies, and identified 250 genes that are transcribed significantly more or less in cancer cells. Molecular biologists zeroed in on the most informative 21 genes, and then bioinformaticians developed an algorithm to quantify recurrence risk based on expression patterns. Next came validation studies. What really sped up the development was testing the gene-expression panel in tumour tissue stored in paraffin. "Instead of a prospective trial that would have taken 10 to 15 years, we could work with studies done in the past with relevance to practice today," says Shak.

Genetic profiling in cancer, of mutations as well as expression patterns, is done in academic laboratories as well as at pharmaceutical, biotech and genetic testing companies. Traditional training in cancer biology using any model system is good preparation, and researchers





Cancer combatants: Huijun Ring (top); reviewing a test slide (bottom); and kidney cancer cells (left).

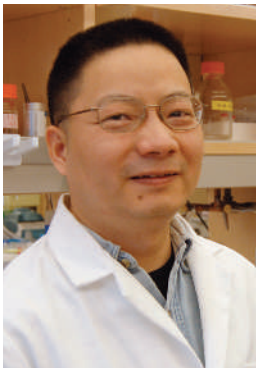
can make the transition into working with human tissues or people. Complementary expertise, such as clinical-trial design, can be acquired in postdoctoral research positions or on the job.

Dual satisfaction

Zhenghe John Wang, a geneticist at the Case Comprehensive Cancer Center in Cleveland, Ohio, stayed in the academic arena. He did his graduate work using yeast, a classic system for studying cell-cycle genes. He then did a postdoc with colon-cancer guru Bert Vogelstein at Johns Hopkins University, and Wang's group at Case has recently identified a new pathway in that cancer. Wang says he has the dual satisfaction of conducting exciting research and helping patients. "Academic cancer research is more hypothesis-driven, but the differences between pharmaceutical and academic research are not as great as they once were," he says.

At Genomic Health, the focus is on development more than discovery. Graduates in biochemistry, chemistry, molecular biology or pharmacology who welcome the challenge of developing precise assays are in great demand, says Shak. Many researchers come from the pharmaceutical and biotech industries, with a background in drug development or developing assays.

An alternative to on-the-job training in drug development or clinical-trial design is a specialized postdoc straddling academia and business. This was the path of Huijun Ring, director of pharmacogenetics at DNAdirect in San Francisco. DNAdirect offers genetic tests for breast, ovarian and colon cancers. A test for breast cancer detects a single gene variant that predicts response to the drug tamoxifen; a test for colon cancer looks for 23 mutations.



Ari Melnick (top) and Zhenghe John Wang.

CHUNGUANG GUO

Ring earned a PhD in molecular biology and genetics at Cornell University working with fruitflies. "As a postdoc I wanted to work more closely with a project impacting human health," she says. After studying clinical molecular genetics at Stanford University, Ring put that expertise to use at InCyte, in Wilmington, Delaware. As head of pharmacogenomics, she helped drug companies develop products and services including clinical-trial design. At DNAdirect she works with patients and doctors to use clinical tests. Skills include communication, project management and understanding how science can help people, she adds.

Perhaps more than any other biomedical field, the new, more complex view of cancer requires savvy with statistics. Biomedical scientists hoping to use genomic technology in predictive medicine need training in biostatistics, says Richard Simon, chief of the biometric research branch at the US National Cancer Institute. "Computational training is also useful, but less essential," he says. Errors in data analysis are so common, notes Simon, that he urges young scientists to go beyond the one required statistics course in their training.

Software Simon developed to analyse gene-expression data was used in the development of MammaPrint, a newly approved 70-gene signature for breast cancer that was developed at the Netherlands Cancer Institute in Amsterdam and its offshoot Agendia, and has been validated in two multinational studies. An opening at Agendia for a bioinformatician illustrates the value of technical and computational skills: it calls for a PhD or a master's with two years of experience that includes using microarrays to analyse tumour samples and work with "complex clinical and molecular databases".

But Ari Melnick, assistant professor of developmental biology and molecular biology and medicine at the Albert Einstein College of Medicine of Yeshiva University in New York City, points out that it is simply impossible for any one person to have expertise in the various areas required to take gene-expression profiles from discovery to the clinic. "This research involves an interface among computational biologists, molecular biologists and biostatisticians," he says. His team developed a 485-gene-expression test that predicts response to a particular drug candidate in patients with B-cell lymphoma.

Going translational

Because genetic analysis in cancer lies at the crossroads of translational ('bench to bedside') research, scientist–physician teams are critical. In the haematology lab at Brigham and Women's Hospital, Boston, there are as many PhDs as MDs. Postdoc Stefan Fröhling, an MD, calls this ideal. Physicians, he says, know the clinical challenges and often have access to collections of well-annotated tumour samples. PhDs know the biology and have the experimental skills to study the functional implications of molecular genetic abnormalities.

Developing genetic tests is much more specific than it was a decade ago, but the work blurs boundaries of expertise. "It is exciting to watch people integrate," says Rick Lesniewski, director of oncology research at Abbott. "A geneticist might talk to a crystallographer, who talks to a medicinal chemist, who talks to a pharmacologist. Seeing how disciplines come together to generate progress in the evolution of a drug is amazing." ■

GENOMIC HEALTH

Ricki Lewis is a freelance science writer in Schenectady, New York.

NATURE|Vol 446|12 April 2007

MOVERS

Stephen Cohen, director, Temasek Life Sciences Laboratory, Singapore

1996–present: Coordinator, developmental-biology unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany
1993–present: Group leader, EMBL, Heidelberg, Germany
1990–93: Assistant investigator, Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas

Geneticist Stephen Cohen eagerly anticipates taking his experience to yet another continent. Following stints in Europe and North America, Cohen will be heading to Asia as director of the non-profit Temasek Life Sciences Laboratory (TLL). During visits to Singapore, Cohen has been impressed with the working style of the city-state's growing science community. "I was very taken with the energy, excitement and enthusiasm of science in Asia, especially Singapore," he says.

Cohen goes to the five-year-old institute after serving as coordinator of the developmental-biology unit at the European Molecular Biology Laboratory (EMBL) in Germany. The Canadian geneticist has long been interested in developmental biology and studying how cells form patterns to organize an animal body. After earning his PhD in biology at Princeton University in New Jersey in 1983, he worked for three years as a postdoctoral fellow at Massachusetts Institute of Technology and the nearby Whitehead Institute. After another postdoc stint at the Max Planck Institute for Developmental Biology in Germany, he became an assistant professor and Howard Hughes Medical Institute assistant investigator at Baylor College of Medicine, Houston.

In 1993, Cohen become a group leader of EMBL and in 1996 the head of the newly formed developmental-biology unit. In addition to studying morphogen gradients and pattern formation, his team studies micro-RNAs and has identified genes that influence the control of metabolism and tissue growth. Cohen says he learned a lot from EMBL, which requires labs to be small, making frequent interdisciplinary interactions critical. "I hope to bring some of these approaches to TLL," he says. He also plans to bring his lab members to Singapore to continue his work.

Cohen will probably be expected to play a different role from that of the previous director, William Chia, who nurtured the fledgling institute by recruiting a number of high-quality scientists (see *Nature* 427, 660; 2004). Chia says he achieved his goal of assembling a multidisciplinary group of 25 labs and 210 researchers. These researchers, many in their early thirties, appreciate the atmosphere of transparency and collaboration, says Chia, as well as the energy of the place.

"TLL is a young, exciting place," says Cohen. "It will be my job to help it reach a high level of international recognition as the lab matures."
Ichiko Fuyuno

NETWORKS & SUPPORT

Crossing boundaries

Throughout my career I have always tried to resist the almost irreversible process of specialization and disciplinary parochialism triggered by the expansion of knowledge. A fortuitous encounter with a novel sort of fellowship has helped me do so. It's also helped me reflect on the impact that my work could have on society.

A graduate of University College London, I pursued my doctoral studies at the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany. During those years, I recognized the need to frame the proceedings of science and their impact as part of a larger context. I pursued this aim by becoming a member of the EMBL Science and Society committee. When I began to search for postdoc opportunities, I looked for a formal way to embrace transdisciplinarity and to bridge science and society. Unfortunately, few funding organizations offered opportunities at my career stage.

Then, in 2003, Swiss entrepreneur Branco Weiss set up an innovative Society in Science fellowship to explore new avenues in the relationship between the two. It is aimed primarily at life-science postdocs seeking to incorporate novel social and cultural aspects into their work. Starting with a rigorous scientific question related to their field, fellows are given the intellectual freedom to develop a sound transdisciplinary research agenda incorporating perspectives from areas such as the history and philosophy of science, sociology, anthropology, public health and law. Fellows receive support for up to five years and carry out projects at one or more institutions of their choice.

With this fellowship, I am now conducting research on behavioural neuroscience while carrying out sociological studies of the impact such experimentation has on society, with special focus on the growing use of psychopharmacological drugs. This has helped me be more critical of my own work and interact with scholars outside my own field.

Natural scientists often debate with social scientists or humanities scholars on important societal issues. But they rarely have much knowledge of each other's jargon and methods. This type of postdoctoral research offers the opportunity to merge two worlds into one unique experience — and to widen the prospects for a more stimulating career.

Giovanni Frazzetto is a Branco Weiss fellow at EMBL in Monterotondo, Italy, and at the BIOS Centre of the London School of Economics.

POSTDOC JOURNAL

Missing the mundane

I'm coming to the conclusion that the real challenge with moving departments is not coping with the big changes, but dealing with all the little ones. I had expected, and prepared for, the fact that I'd be working for different people on a different project. As making those changes is often the reason you're moving in the first place, you embrace them.

It's the minutiae that get to me. It's hard for me to feel at home when I get lost every time I venture out onto campus, or when I don't know how the printers and photocopiers work. In my first week here I spent the better part of a day running around to three different offices (on opposite sides of the campus, naturally) trying to get registered. I'd yet to find that one person in the maze of bureaucracy who actually knows what's going on.

It's the same outside work, especially as I have moved to a new country. I'm having fun exploring, but most of the beer is cold and fizzy, I've yet to locate a reliable source of proper English cheddar, and I'm still auditioning for my new coffee haunt.

Knowing all those little things is what makes you feel that you are living in a place, rather than just visiting. I just hope that I get time to discover them all before I have to leave.

Chris Rowan is a postdoctoral student in the geology department at the University of Johannesburg, South Africa.

The inside track from academia and industry

Tough sell

Let's talk money: what you're likely to earn, how to get a pay rise and when to consider other options.

Martin Lang

Salaries in the science industry have long been a topic of discussion. What sort of pay should scientists expect at various stages of their careers — as recent graduates, postdocs and scientists with some experience in the workplace? How should they approach pay negotiations? How much should an employer expect to spend on salaries?

Many recent graduates have no idea how much they should expect to earn. Where should they look for actual statistics? Talking to friends and former colleagues is one possibility. Studying statistical data delivered by institutes or professional institutions is another. But the situation is more complicated than it may seem. Salaries depend on a variety of issues, including company size, national or regional location and, of course, individual qualifications, skills and work experience. And can you trust everything you read? It's important to remember that these statistics often come from surveys. People are more likely to fib about salaries than just about any other subject.

Smart young academics applying for an entry-level position often have high hopes. With their master's or PhD, perhaps from a prestigious university, they may consider themselves to be among the élite of their nation. They have survived on what seems like modest pay at universities or public-service institutions, and they expect better. But academic and public-service postdoc positions actually pay rather well, compared with entry-level jobs in industry. Work experience does help, though, and pay rises tend to be larger and more frequent than at universities.

"The desire for more pay should not be the primary motivation for seeking a new job."

When we interview young academics, they often seem to have unrealistic expectations about their first real salary. They stare blankly when we try to explain what's on offer in the real world. They are surprised to learn that medium-sized companies or small research organizations will not fulfil their financial expectations. Only big companies can pay those sizeable entry-level salaries. But openings there are rare, and the companies can choose from a large number of qualified candidates.

Young academics should try to strike a happy medium. They should first of all look at the company (does it have a good reputation?) and at the job (is it really the one for me?). Only as a next step should they think about the money. Find a job you love, and the bonus is that you are likely to do well and be in line for promotions or pay rises.

What about scientists with some employment experience? First of all, the desire for more pay should not be the primary motivation for seeking a new job. If you're otherwise contented, speak to your manager. Make your case for a pay rise on the basis of your performance: an argument based solely on living costs is not likely to win you more money. Argue on the basis of your skill set and the extra projects you have taken on for the company. Your success at negotiating could depend on your personality, qualifications and efforts within the company. Always keep in mind that you have the option of asking for non-cash benefits such as a company car, laptop or a mobile phone rather than requesting higher wages.

If you are successful, a higher salary may bring with it a change in the scope of your duties and more responsibility. Consider not only what you have done for your company but what kind of new duties you are willing to assume. Ask yourself if you are willing and able to take on new functions: do you really have the time and the appropriate qualifications? You should also consider that you might take away responsibilities from your colleagues. It's not a good idea to cause ill-feeling in the company.

"When moving on, try to evaluate your own qualifications in a realistic way."

When moving on, try to evaluate your own qualifications in a realistic way, in the context of your particular business sector and function. For example, positions in R&D or sales and business development are usually better paid than those in administration. In clinical research, employees of contract research organizations are paid less than their colleagues at sponsoring drug companies.

From the perspective of the hiring company, the competition for highly qualified scientific staff has grown in recent years and will keep on growing. Those companies that do not pay as well as or better than their competitors could lose specialists. Moreover, scientists often put their heart and soul into their job, but rarely get the appropriate financial reward. Undervaluing employees often leads to staff attrition. As an employer, it's important — now more than ever — to be aware of changes in the job scene and shifting moods among your scientific staff.

Like any employee, scientists should know how to behave in pay negotiations, how to prepare and how to argue. And remember that accepting lower pay than you wanted does not necessarily mean taking a step backwards in your career. ■

Martin Lang is recruitment consultant for Kelly Scientific Resources in Cologne, Germany.